

IASSIST QUARTERLY

Volume 10 Number 2

Summer 1986

FEATURES

- 3** **Techniques for secondary analysis: Unfolding analysis of "PICK K/N" and "PICK ANY/N" data**
By Wijbrandt H. van Schuur
- 19** **Public Data Use: A View From The Telecommunications Industry In The United States**
By Dianne Schmidley
- 23** **Attrition and the National Longitudinal Surveys of Labor Market Experience: Avoidance, Control and Correction**
By Patricia Rhoton
- 36** **CAST: Centre for Applications Software and Technology**
By A. Stacey
- 41** **Australian Health Statistics**
By Roger Jones
- 46** **Providing Local Data Services**
By R. de Vries
- 48** **The Development of a Canadian Union List of Machine Readable Data Files (CULDAT)**
By Edward Hanis

DEPARTMENTS

- 51** **SES Archiving Policy**
- 54** **APDU Conference**
- 55** **USIA**
- 56** **Archive Course**
- 57** **HRAF News**
- 58** **IASSIST Constitution**
- 46** **Providing Local Data Services**

IASSIST

QUARTERLY

Volume 10 Number 2

Summer 1986

FEATURES

- 3** Techniques for secondary analysis: Unfolding analysis of "PICK K/N" and "PICK ANY/N" data

By Wijbrandt H. van Schuur

- 19** Public Data Use: A View From The Telecommunications Industry In The United States

By Dianne Schmidley

- 23** Attrition and the National Longitudinal Surveys of Labor Market Experience: Avoidance, Control and Correction

By Patricia Rhoton

- 36** CAST: Centre for Applications Software and Technology

By A. Stacey

- 41** Australian Health Statistics

By Roger Jones

- 46** Providing Local Data Services

By R. de Vries

- 48** The Development of a Canadian Union List of Machine Readable Data Files (CULDAT)

By Edward Hanis

DEPARTMENTS

- 51** SES Archiving Policy

- 54** APDU Conference

- 55** USIA

- 56** Archive Course

- 57** HRAF News

- 58** IASSIST Constitution

- 46** Providing Local Data Services

NEW INDIVIDUAL RATES
FOR
Paradigm Press
PUBLICATIONS

SCOPE

The only newsletter reporting regularly on computer applications in the humanities and social sciences. Regular updating on hardware, software, courseware, grants, data bases, networks, and publications. A comprehensive informational calendar.

Computers
AND THE
Humanities

The premier journal in its field, established in 1966. Refereed scholarly articles on research applications in literature, language, history musicology, cultural anthropology, and archaeology. Comprehensive book reviews and review essays. Reviews of software, hardware, and courseware.

Computers
and the
Social Sciences

The new refereed scholarly journal covering computer applications in sociology, anthropology, political science, urban studies, psychology, and related fields. Comprehensive book reviews and review essays. Reviews of software, hardware, and courseware.

Each of these publications is now priced at only \$25.00 for *individuals paying by personal check*. To enter your subscription, return the reply card below.

Please enter my individual subscription(s) to the publications checked. Payment will be by *personal check*.

- | | |
|--|---|
| <input type="checkbox"/> SCOPE | <input type="checkbox"/> My check is enclosed. Please begin |
| <input type="checkbox"/> CHum | my subscription immediately. |
| <input type="checkbox"/> CaSS | <input type="checkbox"/> Please send an invoice. |
|
<input type="checkbox"/> I want a copy of <i>Data Bases in the Humanities and Social Sciences</i> at \$39 plus \$2 domestic postage. | |

Name _____

Address _____

Editorial Information

The IASSIST QUARTERLY represents an international cooperative effort on the part of individuals managing, operating, or using machine readable data archives, data libraries, and data services. The QUARTERLY reports on activities related to the production, acquisition, preservation, processing, distribution, and use of machine readable data carried out by its members and others in the international social science community. Your contributions and suggestions for topics of interest are welcomed. The views set forth by authors of articles contained in this publication are not necessarily those of IASSIST.

Information for Authors

The QUARTERLY is published four times per year. Articles and other information should be typewritten and double-spaced. Each page of the manuscript should be numbered. The first page should contain the article title, author's name, affiliation, address to which correspondence may be sent, and telephone number. Footnotes and bibliographic citations should be consistent in style, preferably following a standard authority such as the University of Chicago press *Manual of Style* or Kate L. Turabian's *Manual for Writers*. Where appropriate, machine-readable data files should be cited with bibliographic citations consistent in style with Dodd, Sue A. Bibliographic references for numeric social science data files: suggested guidelines. *Journal of the American Society for Information Science* 30(2):77-82, March 1979. If the contribution is an announcement of a conference, training session, or the like, the text should include a mailing address and a telephone number for the director of the event or for the organization sponsoring the event. Book notices and reviews should not exceed two double-spaced pages. Deadlines for submitting articles are six weeks before publication. Manuscripts should be sent in duplicate to the Editor:

Walter Piovesan
Research Data Library
W.A.C. Bennett Library
Simon Fraser University
Burnaby, B.C., V5A 1S6 CANADA
(01)604/291-4349 E-Mail: Piovesan@SFU.MAILNET

Book reviews should be submitted in duplicate to the Book Review Editor:

Kathleen M. Heim
School of Library and Information Science
Louisiana State University
Coates Hall, Room 267
Baton Rouge, Louisiana 70803 USA
(01)504/388-3158

Key Title: Newsletter - International Association for Social
Science Information -Service and Technology

ISSN - United States: 0739-1137 Copyright © 1985 by IASSIST. All rights reserved.

Digitized by the Internet Archive
in 2010 with funding from
University of North Carolina at Chapel Hill

Techniques for secondary analysis: Unfolding analysis of "PICK K/N" and "PICK ANY/N" data

by Wijbrandt H. van Schuur¹
Faculty of Social Sciences
University of Groningen
Oude Boteringestraat 23
9712 GC Groningen
The Netherlands

Introduction

In survey research we regularly encounter the following type of question: "which of these stimuli do you prefer most? which of the remaining ones do you now prefer most?" etcetera. Sometimes a full rank order of preferences is asked in this way, more often only a partial rank order is obtained.

¹Paper prepared for presentation at the IFDO/IASSIST Conference, Workshop on Techniques for Secondary Analysis, Amsterdam, May 20 - 23, 1985. *Comments are welcomed by the author.*

Sometimes the question asked is only: "which k of these n stimuli do you prefer most?" or, even more generally, "which of these n stimuli do you prefer?" Such questions can be referred to as 'rank n/n', 'rank k/n', 'pick k/n' and 'pick any/n' data, respectively. Stimuli may be political parties, candidates, career possibilities, or brand names of some consumer good. Rather than asking about 'preference', the questions may also be phrased in terms of other evaluative concepts, such as 'sympathy' or 'importance'. In this paper I will be concerned with analyzing data of the form 'pick k/n' or 'pick any/n'.

Generally these data types are difficult to analyze. Often responses to such data are only reported in the form of frequency distributions of the number of times a stimulus is mentioned as most preferred, second most preferred, etcetera. Trying to find structure in these responses with the help of standard techniques, such as factor analysis or cumulative scaling, is not possible either because no full set of responses to all stimuli is available, or because the responses given are not independent. It is then difficult to determine whether or not all responses given were based on the same underlying criterion. In this paper an analysis technique is presented that allows one to look for structure in the responses to 'pick k/n' or 'pick any/n' questions. Since complete or partial rank orders can always be recoded to the 'pick k/n' form, and since survey questions with independent responses, such as five-point Likert items, can be recoded to the 'pick any/n' form, the type of data analysis presented here can have very general application.

The data analysis technique presented here is a dichotomous version of the unfolding model, proposed by Coombs (1950, 1964), as 'parallelogram analysis'. It differs from Coombs' original proposal in the following ways: the technique proposed here allows for some error (i.e., it conforms to a stochastic model), and it is an exploratory technique to search for

maximal subsets of stimuli that can be represented in a unidimensional unfolding scale. In both these aspects the 'parallelogram analysis' model proposed here resembles the stochastic unidimensional cumulative scaling technique developed by Mokken (1971). The reader should be warned that the technique presented here is not an all-purpose technique for analyzing 'pick k/n' or 'pick any/n' data, but only for those types of data which can be expected to conform to the unfolding model!

The perfect unidimensional unfolding model for complete rank orders of preference

In this section I will first summarize some basic ideas behind unfolding analysis, by using an example from Meerling (1981). In an investigation by Ritzen and Van de Kloot, preference rank orders were collected for the following statements:

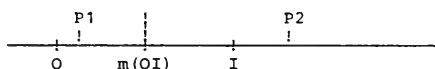
- O : People can be changed in any conceivable direction, provided that the environment is manipulated in the proper way (O = omgeving, environment);
- I : The major condition for people to change is for them to have a clear understanding of their situation (I = inzicht, understanding);
- E : Behaviour is determined much more strongly by emotions than by rational considerations (E = emoties, emotions);
- A : Inborn characteristics determine to a large extent what kind of person someone becomes (A = aangeboren, inborn).

These four statements were shown to psychologist colleagues, and the following six types of preference rank orders were found: OIEA, IOEA, IEOA, EIOA, EAIO, and AEIO.

In applying the unfolding model we assume that there is a latent dimension on which each of these statements can be represented. Meerling suggests for these statements and these preference rank orders that a 'nurture-nature' dimension may be appropriate, in which the statements are arranged in the order O I E A.

When the location of each of the statements on this dimension is established, the dimension can be divided into two areas for each pair of statements (I,J): the first area, in which the first statement is preferred over the second, and the second area, in which the second statement is preferred over the first. The boundary between these two areas lies in the middle between these two stimuli, and is called the 'midpoint of the pair of stimuli', $m(IJ)$. This midpoint allows us to locate individuals who give their preference rank order along this dimension. An individual, P1, who prefers statement O to statement I will be located to the left of midpoint $m(OI)$, whereas another individual, P2, who prefers statement I to statement O, will be located to the right of that midpoint. (see Figure 1)

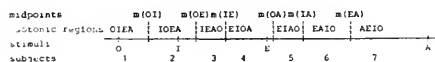
Figure 1
Midpoint $m(OI)$ divides the dimension into two areas



The four statements, together, have six midpoints. These divide the dimension into seven areas, the areas that are separated by the midpoints. Each of these areas is characterized by a special preference rank order, and is called an 'isotonic region'. (see figure 2)

Figure 2

4 stimuli, 6 midpoints, and 7 isotonic areas or subject types



A subject is usually represented on the scale by a single point, called his 'ideal point'. The preference order of the subject is called his 'Individual scale', or 'I-scale' for short. The representation of all subjects and all stimuli jointly on the same dimension is called the 'Joint-scale', or 'J-scale' for short. The I-scale gives the order of the stimuli in terms of their distance from the ideal point of the individual. In other words: the I-scale has to be 'unfolded' at the ideal points to produce the J-scale.

Unfolding analysis is designed to find a joint representation of stimuli and subjects in one dimension, that is, to find a unidimensional J-scale on the basis of the preference rank orders of the individual I-scales. Finding a J-scale brings us two things. The first is an unfoldable order of the stimuli which can generally be used to infer the criterion used by the subjects in determining their preference order (e.g., the nurture-nature criterion). Secondly, having a J-scale allows us to combine a subject's answers to the n survey questions in a single rank order which can be used to measure the preference of the subject in terms of his ideal point on the criterion dimension. By measuring a subject's preference in this way we can create a new variable which can be related to other characteristics of the subject. The purpose of creating such a new variable is to try to explain why people differ in their preferences, or to explain other attitudes or behaviours on the basis of scale values on the

J-scale.

If we have perfect data, such as we usually find in textbooks on scaling (and by perfect data I mean I-scales that can be perfectly represented in a unidimensional unfolding scale) it is no problem to find the J-scale that represents the I-scales. Problems only arise when the data are not perfect, which is in most cases. The major reason why the unfolding technique has so far been relatively unpopular and why it has as yet not been incorporated into most standard statistical packages, is that up to now we have not been able to unfold imperfect data in a satisfactory way. If we could find a usable unfolding technique, interest in it should be great, since the model is plausible, and there is a great deal of interest in measuring the preferences of subjects.

Discussion of some alternative proposals for unfolding models

Before introducing my own model, I will first consider five strategies that have been developed in the literature and which attempt to find useful and interpretable unfolding results. These strategies are all derived from a description of the ideal type of unfolding analysis, namely the perfect representation of a complete rank order of preferences in a unidimensional space, in which all stimuli and all individuals can be represented. These five strategies are:

1. Analyze the I-scales after they have been dichotomized into the k most preferred and $n-k$ least preferred stimuli;
2. Relax the criterion of perfect representation to allow stochastic representation;
3. Find a representation in more than one dimension;

4. Find a representation for a maximal subset of the stimuli;
5. Find a representation for a maximal subset of the subjects.

The first strategy is to dichotomize full or partial rank orders of stimuli into the k most preferred and the $n-k$ least preferred stimuli. The unfolding analysis of such data, parallelogram analysis, can be defended with the argument that the stimuli a subject prefers most will be the most salient ones for him, and a subject will therefore be able to single them out more reliably than the remaining ones. Moreover, although the unfolding model assumes that successively chosen stimuli are in a sense substitutes for the subjects' most preferred stimulus to a decreasing degree, gradually, in the course of giving a full rank order of preference, a subject may begin to use other criteria. Coombs (1964) talked about the 'portfolio model' in this respect, and Tversky (1972, 1979) suggested an 'Elimination by Aspects' model, in which different criteria for preference are hierarchically ordered. If we are interested in finding the dominant criterion that is used first by all respondents, then we should restrict ourselves to analyzing only the first few most preferred stimuli, lest we run the risk of introducing idiosyncratic noise.

Two more practical advantages of this strategy can be mentioned. First, if applying an unfolding model in which the distinction between the k most preferred and $n-k$ least preferred stimuli does not lead to a good-fitting representation, it is no use trying more sophisticated models that require the full rank order, or that may even require metric preference information. Second, the unfolding of dichotomous data implies that essentially all types of data can be used in a preference analysis, as long as the most preferred responses can be distinguished from the others.

The second strategy is to relax the criterion of perfect representation to allow stochastic representation. I regard it as obvious that preference judgments reflect so many idiosyncratic influences, that we should be happy to find that a rather heterogeneous group of subjects agrees on at least a dominant criterion. Stochastic models have been proposed before (Sixtl, 1973; Zinnes and Griggs, 1974; Bechtel, 1976; Jansen, 1981). I regard these proposals as inferior to the model I propose for at least two reasons. Firstly, many of the probabilistic unfolding models assume that the order of stimuli along the J-scale is already known, and only parameter estimation of subjects and stimuli on the basis of the known order is needed. In many cases such an approach is begging the question, as often the order of the stimuli is not known in advance. Secondly, other stochastic unfolding models require that for each subject, we need the probability of his preferring one stimulus to another. In many practical applications this information is impossible to obtain: it is expensive and time consuming enough to ask respondents one single time to compare all pairs of stimuli with respect to preference.

A third strategy to analyze data that are not unfoldable in one dimension is to try to represent them in more than one dimension. It is possible that subjects did not use a single criterion in making their preference judgments, they may instead have used two or three criteria simultaneously. Multidimensional models have been proposed by Bennett and Hays (1960), Roskam (1968), Schönemann (1970), Carroll (1972), Young (1972), Gold (1973), Kruskal et al (1973), and Heiser (1981), among others. They are appealing, because the use of more than one dimension implies the possibility of using a number of additional models that differ in the way in which the various dimensions are combined: the vector model, the weighted distance model, or the compensatory distance model, to mention only a few.

There are at least four possible problems with the multidimensional unfolding model. First, in applying a nonmetric multidimensional unfolding model, we may find an almost degenerate solution, in which most subjects are close together in the centroid of the space, and most stimuli lie in a circle around it. Secondly, also with respect to nonmetric multidimensional unfolding, there is a fundamental difference between the nonmetric analysis of similarities data and the nonmetric analysis of preference data, even though both models are based on the same principle. In the multidimensional analysis of similarities, the isotonic region in which a stimulus falls becomes so small that for a sufficient number of stimuli each stimulus can only be represented by a point in the space, rather than by a region. But in multidimensional unfolding, the representation of some respondents in the form of such isotonic regions is different; some isotonic regions do not shrink to points, but remain open. Such respondents cannot be uniquely represented by one point in the space. Thirdly, multidimensional unfolding assumes that all dimensions are used simultaneously, rather than in a hierarchical order. This is an empirical question, rather than an untestable assumption. Fourth, the assumption that all dimensions are appropriate for all stimuli is equally an empirical question, rather than an untestable assumption.

We are told that reality is not unidimensional. Indeed, a chair has a colour, a weight, and a number of sizes. A person has an age, a sex, and a preference for certain drinks. And a political party may be large, religious and right wing. Still, we never analyze reality. We analyse aspects of reality! We do not compare chairs, subjects or political parties, but sizes of chairs, ages of subjects and ideological positions of political parties. That objects or subjects have more aspects than the ones in which we are interested, does not at all imply that our analyses need to be multidimensional. They may be, but that is an empirical question, and

not an untestable assumption from the outset. I do not fundamentally object to a multidimensional representation of the preferences of a group of subjects. There may be instances in which this is indeed the best model. But the utility of different models will have to be shown in their practical applicability.

With respect to the last two strategies for salvaging the unfolding model, selecting a maximal subset of stimuli and selecting a maximal subset of subjects, it is established practice in multidimensional unfolding analysis to assign stress values to subjects. This implies that any difficulties in finding a representation can be explained by pointing at subjects who used different criteria, or who perhaps even behaved completely at random. A possible procedure, given this assumption, is to delete respondents whose stress values are too high.

However, it may be the case that large stress values occur because one or more stimuli cannot be represented since they do not belong in the same universe of content as the other stimuli. Subjects are allowed to differ in their evaluation of the stimuli, but for unfolding to be applicable, they must agree on the cognitive aspects of the stimuli; whether gentlemen prefer blondes or brunettes is a different matter from establishing whether Marilyn is blonde or brunette. If there is no agreement among the subjects on the characteristics of a stimulus, differences in preference will be difficult to represent.

Often, subjects are selected as representatives of a larger population. Deleting subjects lowers the possibility of generalizing from a sample to a population. Stimuli, on the other hand, are often not so much a random sample of a population of stimuli, but are more often intended to serve as the best and most prototypical indicators of a latent trait; we are often not so much interested in the actual stimuli, but rather in their implications for measuring subjects along this latent trait. This

means that we generally can delete stimuli with less harm than when we delete subjects.

The discussion of these strategies is intended to justify the strategy adopted in the technique to be described below, of finding a stochastic representation of a maximal subset of stimuli and all subjects in one dimension, using the first few preferences of each subject.

Unfolding dichotomous data: the concept of 'error'

We generally do not know in advance which stimuli can be represented in an unfolding scale, nor in which order they can be represented. The approach used here is a form of hierarchical cluster analysis, in which first the best, smallest unfolding scale is found, and then is extended by more stimuli, as long as they continue to satisfy the criteria of an unfolding scale. The smallest unfolding scale consists of three stimuli, since it takes at least three stimuli to falsify the unfolding model. If stimuli A, B, and C form a perfect unfolding scale in this order, then subjects who prefer A and C but not B, do not exist. For the unfolding scale ABC the response patterns in which A and C are preferred but B is not, is defined as the 'error pattern' of that triple of stimuli. But since we do not know in advance in what order the stimuli form an unfolding scale, we must take into account the three permutations in which each of the three stimuli is the middle one: BAC, ABC, and ACB. If a subject prefers A and B, but not C, for example, he makes an error according to the unfolding scale ACB.

For each triple of stimuli, given a dichotomous response to each stimulus, eight response patterns are possible: 111, 110, 101, 011, 100, 010, 001, and 000. If these stimuli form part of an unfolding scale, then one of these eight

patterns cannot occur: the pattern '101' (see Table 1)². This pattern is called the 'error response'.

For each triple of stimuli, in each of its three possible permutations, the frequency of occurrence of the error pattern can be counted. Counting frequencies of error response patterns can be extended to larger response patterns, in which each subject evaluates more stimuli. Table 2 gives five response patterns in which two or three stimuli are preferred from a set of four. In the first two response patterns only one triple is in error. In the last three response patterns two triples are in error. The amount of error in a response pattern is defined as the number of triples in that response pattern that are in error. The last three response patterns therefore contain twice as much error as the first two.

In the second example, four subjects prefer six out of seven stimuli. It makes an enormous difference to the amount of error in their response patterns whether the stimulus not preferred is D, C, B, or A. In the case of D, the amount of error is maximal, whereas in the case of A there are no errors at all.

Stochastic unfolding

The stochastic aspect of the unfolding strategy proposed here lies in comparing the amount of error observed with the amount of error expected under statistical independence. In the deterministic unfolding model, the k stimuli that are preferred by a subject are found within the symmetric closed interval around the subject's ideal point. The probability of preferring a set of stimuli (e.g., two, three, or more) will be '1' if all stimuli fall within the subject's preference

² Editor's note: Tables are gathered together at end of article

interval, and '0' if at least one stimulus falls outside this interval.

The null model differs from the deterministic model in two ways. First, local independence is assumed among preference responses for different stimuli. This means that for each subject the probability of a preference response pattern to a set of stimuli is the product of the positive (preferential) response to each of the stimuli. Second, the null model assumes that there are no individual differences in the probabilities of giving a positive preference response to the stimuli. The expected frequency with which a set of stimuli is preferred will therefore be the product of the relative frequencies with which each stimulus is preferred times the number of cases, if subjects are free to select as many 'most preferred' stimuli as they wish:

$$\text{Exp.Freq}(ijk,101) = p(i).(1 - p(j)).p(k).N$$

where $p(i)$ is the relative frequency with which stimulus i is preferred and N is the number of cases.

The expected number of errors under the null model for 'pick k/n ' data is first explained for 'pick $3/n$ ' data. It consists of two steps:

1. determine the expected frequency of the '111' response pattern by applying the n -way simple quasi-independence model (e.g., Bishop et al, 1975);
2. from the '111' responses to each triple, other response patterns - like 110, 101, or 011 - can be deduced.

In a data matrix, in which each of the N subjects picks exactly 3 of n stimuli as most preferred, the relative frequency $p(i)$ with which each stimulus is picked can be found. In the null model, these $p(i)$'s are derived from the addition of the expected frequency of triples

(i,j,k) for all combinations of j and k with a fixed i . This expected frequency of triples ij,k , $a(ijk)$, is the product of the item parameters $f(i)$, $f(j)$, $f(k)$ times a general scaling factor f , without interaction effects: $a(ijk) = f.f(i).f(j).f(k)$. The values of f , and each $f(i)$ are found iteratively. (see Table 3) The details of this procedure are given in Van Schuur (1984).

Once the expected frequency of the '111' pattern of all triples is known, the expected frequency of the other response patterns can be found, given that each subject picked exactly 3 stimuli as most preferred. For example: Consider the situation in which there are five stimuli, A, B, C, D, and E, and each subject chooses three stimuli as most preferred. For the unfolding scale ABC the error response pattern is the pattern 101, in which stimuli A and C are picked, but stimulus B is not. If B was not one of the subject's choices, then D or E must have been. We can therefore calculate the expected frequency across all respondents of the response pattern 101 for the triple ABC by summing the expected '111' responses of the triples ACD and ACE. In general:

$$\text{Exp.Freq.}(ijk,101) = f.f(i).f(k).\sum_{s \neq i,j,k} f(s)$$

This procedure can easily be generalized to the 'pick k/n ' case, where $k = 2$, or where $k > 3$. First, the expected frequency of each k -tuple, ranging between 1 and $\binom{n}{k}$ is found. Second, the expected frequency of the error response pattern of an unfolding scale of three stimuli is found by calculating:

$$\text{Exp.Freq.}(ijk,101) = f.f(i).f(k).Q$$

where Q is the sum over all $\binom{n-3}{k-2}$ $k-2$ tuples of the product of their $f(s)$'s, where s is not equal to i, j , or k .

Once we know the frequency of the error response observed, $\text{Obs.Freq.}(ijk,101)$, as well as the frequency expected under the null model,

Exp.Freq.(ijk,101), for each triple of stimuli in each of its three essentially different permutations, we can compare the two using a scalability coefficient analogous to Loevinger's H (Loevinger, 1948; Mokken, 1971):

$$H_{ijk} = 1 - \frac{\text{Obs.Freq.}(ijk,101)}{\text{Exp.Freq.}(ijk,101)}$$

For each triple of stimuli (ij, and k), three coefficients of scalability can be found: H(ijk), H(ikj), and H(jik). Perfect scalability is defined as H = 1. This means that no error is observed. When H = 0 the amount of error observed is equal to the amount of error expected under statistical independence.

The scalability of an unfolding scale of more than three stimuli can also be evaluated. In this case we can simply calculate the sum of the error responses to all relevant triples of the scale, for both the observed and expected error frequency, and then compare them, using the coefficient of scalability H:

$$H = 1 - \frac{\sum_{\substack{(ijk) \\ (ijk=1)}}^{(P)} \text{Obs.Freq.}(ijk,101)}{\sum_{\substack{(ijk) \\ (ijk=1)}}^{(P)} \text{Exp.Freq.}(ijk,101)}$$

The scalability of single stimuli in the scale can equally be evaluated, by adding up the frequencies of the error patterns observed and expected, respectively, in only those triples that contain the stimulus under consideration, and then comparing these frequencies using the scalability coefficient for each stimulus separately.

MUDFOLD: Multiple Unidimensional unfolding, the search procedure

After having obtained all relevant information about each triple of stimuli in each of its three different permutations (e.g., Obs.Freq.(ijk,101), Exp.Freq.(ijk,101), and H(ijk)), we can begin to construct an unfolding scale. This is a two-step procedure. First, the best elementary scale is found, and second, new stimuli are added, one by one, to the existing scale.

The best triple of stimuli that conforms to the following criteria is the best elementary scale:

1. its scalability value should be positive in only one of its three permutations, and negative in the other two. This guarantees that the best triple has a unique order of representation;
2. its scalability value must be higher than some user specified lower boundary. This guarantees that if the scalability value is positive, it can be given a substantively relevant interpretation.
3. the absolute frequencies of the perfect patterns with at least two of the three stimuli (i.e., 111, 110, and 011) is highest among all triples fulfilling the first two criteria. This guarantees the representativeness of the largest group of respondents.

Once the best elementary scale is found, each of the remaining n-3 stimuli is investigated to determine whether or not it might make the best fourth stimulus. The fourth stimulus (e.g., D) may be added to the three stimuli of the best triple (e.g., ABC) in any one of four places: DABC, AD BC, ABDC, or ABCD, denoted as place 1 through place 4, respectively. The best fourth - or, more generally,

the $p+1$ -st - stimulus must fulfill the following criteria:

1. All new $\binom{p}{2}$ triples, including the $p+1$ -st stimulus and two stimuli from the existing p -stimulus scale, must have a positive $H(ijk)$ -value. This guarantees that all stimuli are homogeneous with respect to the latent dimension.
2. The $p+1$ -st stimulus should be uniquely representable, in only one of the p possible places in the p -stimulus scale. This guarantees the later usefulness and interpretability of the order of the stimuli in the scale.
3. The $H(i)$ -value of the new stimulus, as well as the H -value of the scale as a whole, must be higher than some user-specified lower boundary (see second criterion for the best elementary scale).
4. If more than one stimulus conforms to the criteria mentioned above, that stimulus will be selected which leads to the highest overall scalability value for the scale as a whole.

This procedure, of extending a scale with additional stimuli, can continue as long as the criteria mentioned above are met. If, however, no stimulus conforms to these criteria, the p -stimulus scale is a maximal subset of unfoldable stimuli. A new procedure then starts which begins by selecting the best triple among the remaining $n-p$ stimuli. This procedure, in which, for a given pool of stimuli, more than one maximal subset of unidimensionally unfoldable stimuli can be found, is called 'multiple scaling'.

The dominance and adjacency matrices: visual inspection of model conformity

Once a maximal subset of unfoldable stimuli is found, a final visual check of model conformity can be performed by inspecting the dominance and adjacency matrices. The dominance matrix is a square, asymmetric matrix which contains in its cells (ij) the proportion of respondents who preferred stimulus i but not stimulus j . If the stimuli are in their order along the J -scale, then for each stimulus i the proportions $p(i,j)$ should decrease from the first column toward the diagonal and increase from the diagonal to the last column. The adjacency matrix is a lower triangle that contains in its cells (ij) the proportion of respondents who preferred both i and j . If the stimuli are in their order along the J -scale, then for each stimulus i the proportions $p(i,j)$ should increase from the first column to the diagonal and decrease from the diagonal to the last row. This pattern is called a 'simplex pattern'. Stimuli that disturb these expected characteristic monotonicity patterns should be considered for deletion from the scale. (see Table 4).

Scale values

Once an unfolding scale of a maximal subset of stimuli has been found, scale values for stimuli and subjects must be found. The scale value of a stimulus is defined as its rank number in the unfolding scale. The scale value of a subject is defined as the mean of the scale values of the stimuli that the subject chose as most preferred. Subjects who did not pick any stimulus from the scale cannot be given a scale value, and must be treated as missing data. An example of the assignment of scale values is shown in Table 5.

Respondents may have different response patterns, but be assigned the same scale value. This can be seen by comparing subjects 1, 2, and 3. Subjects 4 and 5 show that a scale value for a subject does not need to be an integer value. Respondent 6 shows that a scale value is assigned to a subject regardless of the amount of error in his response pattern, which in his case is maximal. Subject 7 does not pick any of the 7 stimuli and therefore cannot be represented on this scale.

An example: Pick the 2 most sympathetic of 6 European party groups

As part of the Middle Level Elite Project (e.g., Van Schuur, 1984), sympathy scores for six European party groups in the European Parliament of 1979 were elicited from party activists from 50 political parties in the European Community. The responses of 1786 subjects about their two most sympathetic party groups were analyzed. The six party groups are, with the letter by which they will be denoted, and with the frequency with which they were mentioned as sympathetic in brackets:

A: Communists (359); B: Social Democrats (747); C: European Democrats for Progress (366); D: European Liberals and Democrats (662); E: Christian Democrats (792); and F: Conservatives (646).

The frequency with which each pair of parties was mentioned as most sympathetic is: AB(341) AC(9) AD(3) AE(2) AF(4) BC(106) BD(202) BE(86) BF(12) CD(124) CE(50) CF(77) DE(217) DF(116) EF(437).

On the basis of this information, a labeled matrix can be constructed that contains, for each triple of stimuli in each of its three essentially different permutations, the values

Obs.Freq.(ijk,101), or E(o), Exp.Freq.(ijk,101), or E(e), and H(ijk). This information is given in Table 6.

Table 6 provides all the necessary information for constructing an unfolding scale. First, the best elementary unfolding scale is found among those triples that have a positive scalability value in only one of its three permutations. This leaves the ordered triples ABC, ABD, BCF, BDE, CDE, DCF, CFE, and DEF. Triple ABD is the best triple, since the sums of the pairs (A,B) and (B,D) is highest. The H-value of triple ABD is 0.96, which is well above the recommended default user specification of 0.30.

On the basis of scale ABD, stimulus C cannot be represented in this scale in any position, since the triple B,C,D has negative H-values in all three permutations. Stimulus E is uniquely representable in place 4, forming scale ABDE, whereas stimulus F is representable in either place 1 (scale FABD) or place 4 (scale ABDF). Stimulus E is selected because it is the only one uniquely representable. The four-stimulus scale is ABDE, its H-value is $1 - 93/485 = 0.81$, which is acceptably high. For the best fifth stimulus, we need only consider stimulus F. This is now only representable in place 5, which gives the final scale ABDEF. Its H-value is $1 - 245/1185 = 0.79$.

In the process of scale construction, the H-values of individual stimuli are also calculated. For the triple ABD these values are the same: $H(A) = H(B) = H(D) = H(ABD) = 0.96$. For the four- and five-stimulus scales these values must be computed separately. The resulting H-values for the final scale are shown in Table 7, along with the dominance matrix and the adjacency matrix for the stimuli in the order of the final scale. Neither matrix shows any violation of the expected characteristic monotonicity pattern.

Five of the six European party groups can be included in an unfolding scale based on party

activists' sympathy scores for these party groups. The scale can be interpreted as a left-right dimension, with the Communists represented in the left-most place and the Conservatives in the right-most place. To corroborate this interpretation, I have correlated subjects' scale scores for this unfolding scale with their scores on a left-right self-placement scale. This correlation was 0.66.

The European Democrats for Progress (EDP, stimulus C) was not incorporated in the scale. This party group consists of the French Gaullists (RPR), the largest Irish party Fianna Fail (FF), and the Danish Progress Party (FRP). This party group is not represented in many EC countries, so it is probably less well known than other party groups, and did not, therefore, receive high sympathy scores from respondents who might have been expected to be sympathetic, based on their positions on the scale.

Concluding remarks

The procedure described above for the analysis of 'pick k/n' data can be extended to apply to 'rank k/n' data. Such procedures have been independently proposed by Davison (1978) and by Van Schuur and Molenaar (1982). Using partial rank order information might provide more precise measurements for both the stimuli and the subjects. However, since in this procedure all six permutations of a triple of stimuli have their own observed and expected error patterns, the accuracy of estimation with the same data set decreases sixfold. As table 7 already shows, the $H(ijk)$ -value of some triples is based on a comparison of rather small numbers, and such comparison will therefore be even more difficult in the 'rank'-case. Moreover, for small k the increase in measurement precision is minimal, and I have

already expressed some doubts about the reliability of the k -th preference judgment, when k gets large.

A computer program (MUDFOLD) has been devised to perform a multiple unidimensional unfolding analysis on complete or partial rank order data, 'pick k/n' or 'pick any/n' data, or on the usual attitudinal data, such as Likert items or thermometer scores. The program is interactive, self-explanatory, and very user-friendly. The user may define a startset rather than use the best elementary scale to find a larger unfolding scale, or test the unfoldability of a given set of stimuli in a given order. In either case, if a triple of stimuli in the user defined order has a negative $H(ijk)$ -value, this triple will be flagged, along with its $E(o)$ -, $E(e)$ -, and $H(ijk)$ -values. The output not only consists of the H - and $H(i)$ -values of the final scale, but also gives an overview of which stimuli at which places were candidates for selection at what step of enlargement, and the H - and $H(i)$ -values of the stimuli in the scale at which step of enlargement. Moreover, the output contains a variety of additional information which may help the researcher either find a better scale, or explain why certain stimuli did not fit in the unfolding scale. The computer program is available from the University of Groningen. The development of the unfolding model presented above, together with more than twenty applications, is described in more detail in my dissertation (Van Schuur, 1984).□

References:

- Bechtel, G.G. (1976). Multidimensional preference scaling. The Hague: Mouton.
 Bennett, J.F. and W.L. Hays (1960). Multidimensional unfolding: determining the dimensionality of ranked preference

- data. *Psychometrika*, 25, 27-43.
- Bishop, Y.M.M., S.E. Fienberg, and P.W. Holland (1975). Discrete multivariate analysis: theory and practice. Cambridge, Mass.: MIT Press.
- Carroll, J.D. (1972). Individual differences and multidimensional scaling. In: R.N. Shepard et al (eds.), Multidimensional scaling. Vol. I: theory. New York: Seminar Press.
- Coombs, C.H. (1950). Psychological scaling without a unit of measurement. *Psychological review*, 57, 148-158.
- Coombs, C.H. (1964). A theory of data. New York: Wiley.
- Davison, M.L. (1979). Testing a unidimensional, qualitative unfolding model for attitudinal or developmental data. *Psychometrika*, 44, 179-194.
- Gold, E.M. (1973). Metric unfolding: data requirements for unique solutions and clarifications of Schönemann's algorithm. *Psychometrika*, 38, 555-569.
- Heiser, W.J. (1981) Unfolding analysis of proximity data. University of Leyden: unpublished dissertation.
- Jansen, P.G.W. (1983). Rasch analysis of attitudinal data. Catholic University of Nijmegen - Rijks Psychologische Dienst, Den Haag: unpublished dissertation.
- Kruskal, J.B., F.W. Young, and J.B. Seery (1973). How to use KYST, a very flexible program to do multidimensional scaling and unfolding. Murray Hill: Bell Labs., mimeo.
- Meerling (1981). Methoden en technieken van psychologisch onderzoek, deel 2: data-analyse en psychometrie. Meppel: Boom.
- Mokken, R.J. (1971). A theory and procedure of scale analysis, with application in political research. The Hague: Mouton.
- Roskam, E.E. (1968). Metric analysis of ordinal data. Voorschoten: VAM.
- Schönemann, P.H. (1970). On metric multidimensional unfolding. *Psychometrika*, 35, 349-366.
- Sixtl, F. (1973). Probabilistic unfolding. *Psychometrika*, 38, 235-248.
- Tversky, A. (1972). Elimination by aspects: a theory of choice. *Psychological review*, 79, 281-299.
- Van Schuur, W.H. and I.W. Molenaar (1982). MUDFOLD, multiple stochastic unidimensional unfolding. In: H. Caussinus, P. Ettinger, and R. Thomassone (eds.) COMPSTAT 1982. Part I: proceedings in computational statistics. Vienna: Physica-Verlag, pp. 419-424.
- Van Schuur, H. (1984). Structure in political beliefs, a new model for stochastic unfolding with application to European party activists. Amsterdam: CT Press.
- Young, F.W. (1972). A model for polynomial conjoint analysis algorithms. In: R.N. Shepard et al (eds.), Multidimensional scaling: theory and applications in the behavioral sciences, vol. I: theory. New York: Seminar Press.
- Zinnes, J.L. and R.A. Griggs (1974). Probabilistic multidimensional unfolding analysis. *Psychometrika*, 39, 327-350.

Table 1

Table 1: parallelogram analysis of perfect 'pick 3/11' data

1: subject prefers stimulus

0: subject does not prefer stimulus

```

subjects:      1 2 3 4 5 6 7 8 9
               | | | | | | | | |
stimuli:       A B C D E F G H I J K

```

```

subject nr.  response pattern:
1           1 1 1 0 0 0 0 0 0 0 0
2           0 1 1 1 0 0 0 0 0 0 0
3           1 1 1
4           1 1 1
5           1 1 1
6           1 1 1
7           1 1 1
8           0 0 0 0 0 0 0 1 1 1 0
9           0 0 0 0 0 0 0 0 1 1 1

```

Table 2

Table 2: Two examples of response patterns that contain error

Example 1:			Example 2:		
A B C D	Error in triples		A B C D E F G	Error in triples	
1 0 1 0	ABC		1 1 1 0 1 1 1	ADE ADF ADG BDE BDF BDG CDE CDF CD	
0 1 0 1		BCD	1 1 0 1 1 1 1	ACD ACE ACF ACG BCD BCE BCF BCG	
1 0 0 1	ABD ACD		1 0 1 1 1 1 1	ABC ABD ABE ABF ABG	
1 1 0 1	ACD BCD		0 1 1 1 1 1 1	none	
1 0 1 1	ABC ABD				

Table 3

Table 3: Observed data matrix and matrix with expected frequency

Observed data matrix:

stimuli:										SUM
subjects:	A	B	C	D	...	i	j	k	... n	
1	1	1	1	0		0	0	0	0	3
2	1	1	0	1		0	0	0	0	3
⋮						⋮				
v	0	0	0	0		1	1	1	0	3
⋮						⋮				
N	0	1	0	0		0	1	0	1	3
	p(A)	p(B)	p(C)	p(D)		p(i)	p(j)	p(k)	p(n)	

Matrix with expected frequencies:

stimuli:										
triples:	A	B	C	D	...	i	j	k	...	n
(ABC)	a_{ABC}	a_{ABC}	a_{ABC}	0		0	0	0		a_{ABC}
(ABD)	a_{ABD}	a_{ABD}	0	a_{ABD}		0	0	0		a_{ABD}
⋮						⋮				
(ijk)	0	0	0	0		a_{ijk}	a_{ijk}	a_{ijk}		a_{ijk}
⋮						⋮				
$\binom{n}{3}$										
	$p(A)$	$p(B)$	$p(C)$	$p(D)$		$p(i)$	$p(j)$	$p(k)$		$p(n)$

 a_{ijk} : expected frequency of triple (ijk) $a_{ijk} = f.f(i).f(j).f(k)$ (i.e., no interaction)

The values for f and f(i) are found iteratively.

Table 4

Table 4: Dominance and adjacency matrix for a perfect 4-stimulus unfolding scale

Data matrix					Dominance matrix:				
A	B	C	D	frequency	A	B	C	D	
1	0	0	0	p	A	-	p	p+t	p+t+w
0	1	0	0	q	B	q+u+x	-	q+t	q+t+u+w
0	0	1	0	r	C	r+u+v+x	r+v	-	r+u+w
0	0	0	1	s	D	s+v+x	s+v	s	-
1	1	0	0	t	Adjacency matrix:				
0	1	1	0	u					
0	0	1	1	v					
1	1	1	0	w					
0	1	1	1	x					
					A	B	C	D	
					A	-			
					B	t+w	-		
					C	w	u+w+x	-	
					D	0	x	v+x	-

Table 5

Table 5: Assignment of scale values to stimuli and subjects

stimuli	A	B	C	D	E	F	G	
rank number	1	2	3	4	5	6	7	scale value of subject
subject nr.								
1	0	1	1	1	0	0	0	3
2	0	0	1	0	0	0	0	3
3	1	0	1	0	1	0	0	3
4	0	1	1	0	0	0	0	2.5
5	0	0	0	1	0	1	1	5.67
6	1	1	1	0	1	1	1	4
7	0	0	0	0	0	0	0	- (missing datum)

Table 6

Table 6: Labeled H-matrix for 'pick 2/6' European party groups, N=

	Scale jik			Scale ijk			Scale ikj		
	E(o)	E(e)	H(ijk)	E(o)	E(e)	H(ijk)	E(o)	E(e)	H(ijk)
ABC	106	88	-0.21	9	36	0.75	341	86	-2.97
ABD	202	177	-0.14	3	73	0.96	341	86	-2.97
ABE	86	226	0.62	2	93	0.98	341	86	-2.97
ABF	12	171	0.93	4	71	0.94	341	86	-2.97
ACD	124	75	-0.66	3	73	0.96	9	36	0.75
ACE	50	95	0.48	2	93	0.96	9	36	0.75
ACF	77	72	-0.07	4	71	0.94	9	36	0.75
ADE	217	192	-0.13	2	93	0.98	3	73	0.96
ADF	116	146	0.20	4	71	0.94	3	73	0.96
AEF	437	186	-1.35	4	71	0.94	2	93	0.98
BCD	124	75	-0.66	202	177	-0.14	106	88	-0.21
BCE	50	95	0.48	86	226	0.62	106	88	-0.21
BCF	77	72	-0.07	12	171	0.93	106	88	-0.21
BDE	217	192	-0.13	86	226	0.62	202	177	-0.14
BDF	116	146	0.20	12	171	0.93	202	177	-0.14
BEF	437	186	-1.35	12	171	0.93	86	226	0.62
CDE	217	192	-0.13	50	95	0.48	124	75	-0.66
CDF	116	146	0.20	77	72	-0.07	124	75	-0.66
CEF	437	186	-1.35	77	72	-0.07	50	95	0.48
DEF	437	186	-1.35	116	146	0.20	217	192	-0.13

Table 7

Table 7: Final unfolding scale for 'pick 2/6' European party groups

	p(i)	H(i)
A Communists	0.20	0.96
B Social Democrats	0.42	0.85
D European Liberals and Democrats	0.37	0.71
E Christian Democrats	0.44	0.72
F Conservatives	0.36	0.79
N=1786 H=.79		

Dominance matrix

	A	B	D	E	F
A	-	1	19	19	19
B	17	-	25	31	35
D	30	19	-	18	24
E	41	37	29	-	17
F	32	31	25	7	-

Adjacency matrix

	A	B	D	E	F
A	-				
B	17	-			
D	0	11	-		
E	0	5	12	-	
F	0	1	6	24	-

Public Data Use: A View From The Telecommunications Industry In The United States

by A. Dianne Schmidley,
Staff Manager, Bell Atlantic

Bell Atlantic is one of eight U.S. telecommunications firms resulting from the breakup of the AT&T owned Bell System on January 1, 1984, the largest divestiture and reorganization in corporate history. Bell Atlantic owns an assortment of companies engaged in various aspects of providing telecommunications services and products. These companies can be divided into two categories which we call the "Enterprises Group" and the "Network Services

Group."

The Enterprises Group provides services and products to a variety of geographic locations in the United States and Canada and it is a relatively unregulated entity. Activities of the Network Services Group are concentrated in the seven political jurisdictions of Washington D.C., Delaware, Maryland, New Jersey, Pennsylvania, Virginia, and West Virginia. Management Services, Incorporated (MSI) and the operating telephone companies are part of the Network Services Group.

The Economic Analysis District (EAD), my organization, is located in the Business Planning and Financial Management Department of the MSI. The primary occupation of staff members in the EAD is the provision of internal consulting support to the Network Services Group, which is dominated by the concerns of the operating telephone companies: the Chesapeake and Potomac Telephone Companies of Washington, D.C., Maryland, Virginia, and West Virginia, New Jersey Bell Telephone Company, Bell of Pennsylvania, and Diamond State Telephone Company in Delaware. These concerns can be divided into four functional areas: Regulatory, Personnel, Facilities Planning, and Marketing.

Because the network services group is almost wholly comprised of regulated telephone companies, "regulatory issues" are the most important concern of the EAD. State regulatory agencies, often called Public Utility Commissions, determine, through pricing decisions, who will bear the burden of the rates the companies charge to recoup operating expenses and guarantee the investors in Bell Atlantic stock a competitive rate of return on their investment dollar. Demographic and economic analyses of the size, distribution and composition of each company's market provide the basis for determining the effects of various pricing configurations.

Analytical work undertaken by Bell Atlantic economists and demographers has been made more complicated by the divestiture, since the breakup of the Bell System literally led to the breakup of telephone-served geography. The seven jurisdictions we serve contain 19 "local access and transport areas" or LATAs, which do not correspond to any other political or statistical entity, although they are associated with Metropolitan Statistical Areas (MSAs) in many cases. These LATAs, or large exchange areas, obtain their external communication links from the Interexchange Carriers (IECs), telecommunications companies engaged in long distance calling services which the local telephone companies (such as those owned by Bell Atlantic) are constrained from offering, owing to federal regulatory restrictions. Relations between the IECs and the local telephone companies are regulated by Federal Communications Commission (FCC) rulings, legislative requirements established by the U.S. Congress, and Executive Branch decisions made through entities such as the Justice Department and the Federal Court System. In order to comply with the various rulings and legislative mandates, and sometimes question their logic, Bell Atlantic must have knowledge of the size, distribution and composition of the market within and between the LATAs.

In addition to the Regulatory function, another important activity of the companies the EAD supports is the Personnel function. Whether we are addressing Equal Employment Opportunities issues, employment site location studies, force planning or how to strategically locate our work crews relative to population growth and migration churn, we turn to data available in the public domain to answer questions about the size, distribution and composition of the local labor force and the telephone-served population.

Our third major area of support for the operating telephone companies involves the "Facilities Planning" function. The telephone companies operate networks, which consist of

central offices containing switching systems (large computers), miles of cable, and microwave towers. We are constantly concerned with plant capacity and demand for our services which translates into changes in demand for central office switching and transmission capability. Population and economic forecasts based on public data make the forecasting of demand possible and enhance our ability to plan efficiently and effectively.

Although we are heavily regulated by the government, we do have many marketing concerns, and the market we serve constitutes the fourth area of functional responsibility for the EAD. Some of the more familiar marketing efforts the telephone companies engage in include the distribution of the white and yellow pages directories, and the provision of operator services such as call completion and information retrieval. In addition, we offer products such as business to business directories, and services such as cable television access. We serve the government at the national, state and local level. We serve large industries, such as the steel mills in West Virginia and Pennsylvania, and small enterprises such as a savings and loan company in Maryland. We serve the elderly, the handicapped, homeowners, travellers, and, a new customer since the divestiture, Interexchange Carriers [sic]. Knowledge of the constituents of this market is derived from public data coupled with our own internal surveys.

What are the kinds of public data used by Bell Atlantic? Generally, we use as much of the demographic and economic data as we can obtain from the federal, state and local governments, whether it comes from censuses, surveys or administrative records, but our most important source of demographic or socio-economic information is the 1980 U.S. Census of Population and Housing. These data are available in many forms: published, on microfiche, and on magnetic tapes. The problem is that there is more data than we can handle, so we have implemented an online

demographic data retrieval system to assist us.

As I mentioned earlier, Bell Atlantic has a unique problem. The divestiture left the company with odd service areas called LATAs. In order to provide information to our companies, the EAD modifies public data from the economic and demographic censuses and surveys to make it conform to the geographic area Bell Atlantic serves.

Prior to the divestiture, the operating telephone companies were concerned with the same geographical areas they serve today. The unit of concern, however, was the wire center area or central office district (COD) as it is sometimes called. To complicate matters further, local exchange areas (smaller and different from the LATAs described above) were also a concern. Fifty years ago, all three entities were represented by the same geographic area, corresponding to a community or settlement. Technological change, which allowed the newer central offices to serve more than one of the old wire center areas, population change, and concessions to consumers with regard to their calling access led to an erosion of this one-to-one correspondence. As a result, the telephone companies not only have served and continue to serve areas unlike any other known political or statistical geographical areas, they serve a number of entities that do not correspond to one another.

Because of the continuous need to determine the demographic/economic characteristics of telephone service areas, in order to address the functional areas described above, the requirement for tailored public data arose long before the divestiture. The key to tailoring the demographic and economic data used to develop construction plans, engage in force planning and answer the questions of the regulators is Census geography.

Census tract and block group information, aggregated to user described areas, is the

Rosetta stone of managers engaged in economic and demographic analysis. With the advent, in 1970, of the first fully automated census, the laborious task of aggregating census tract and block group information by hand became, mercifully, obsolete.

Today, there are three major methodological approaches underlying the automated demographic data retrieval systems which provide information for user defined geography.

1. Federal Information Processing Codes (FIPS) are assigned to every political and statistical entity in the United States. This means that all political and statistical geographic units, such as states, counties, MSAs, and census tracts, have unique identification codes. In the automated system, the user can retrieve information associated with these codes. This approach is efficient if the user is seeking information for a list of states, counties, or municipalities. The first attempts to aggregate data for user described areas, such as wire center areas, were based on combinations of block groups/census tracts, and relied on this mechanism. When thousands of geographic units were involved, however, (the old Bell System had 10,000 wire center areas) this particular approach proved to be extremely time consuming, even after automation.
2. The assignment of geo-coordinates (latitudinal/longitudinal coordinate points) to census data provided the basis for a major breakthrough in the automation of demographic data retrieval. Every census block in the United States received, in 1970, a centroid assignment of a unique set of coordinate points. The centroid is the geographic or population center of a block area; there are variations in the way these assignments are made, but discussion of this topic is beyond the scope of this paper. In 1970, point assignments developed by the U.S. Bureau of the Census, were listed in the Master Enumeration Districts List (MEDS) and in 1980, Census Bureau point assignments were listed in

the Master Area File Reference List (MARF). In 1990, they will probably be found in the Topographically Integrated Geographic Referencing and Encoding System (TIGER).

In themselves, the centroid assignments are useless for solving the problem of demographically describing user defined areas. Software linking the coordinate assignments and user described boundaries of study areas, which have been transcribed into binary code, are needed to complete the demographic data retrieval operation. To date, most of the software for this type of application is owned by non-governmental sources, and licensing arrangements must be purchased in order to make use of the private sector product. Before the divestiture, the operating telephone companies in the jurisdictions now served by Bell Atlantic had transcribed their wire center area boundaries into binary coded polygon files. Census data based on the MEDS and MARF assignments could be aggregated to produce demographic profiles of the user described areas. Since LATAs are aggregations of wire center areas, all that had to be done after the creation of the LATAs was to aggregate the wire center polygons into LATA polygons. As was mentioned earlier, LATAs are also aggregations of the smaller exchange areas, and/or COD areas. At the LATA level, however, the difference between the three telephone entities (wire center areas, central office districts, and exchanges) disappears. Thus, aggregating the wire center areas to equate to LATAs does not cause discrepancies. The final result is census data tailored to our LATA areas.

3. The third type of geographic linking system available for tailoring government produced demographic data to meet user defined needs is the Geo-Based Files/Dual Independent Map Encoding or GBF/DIME process. Briefly, this process makes possible the matching of Census address records for urbanized areas with user records. In the case of the telephone company,

these are customer records. Customer records processed through the GBF/DIME program can be linked, at the census tract level, with specific socio-economic characteristics. This process was used to provide the Washington, D.C. Public Utility Commission with information concerning links between telephone availability and characteristics of the inhabitants of areas under study. This process is more limited than the other two, however, since the GBF/DIME files are only available for urbanized areas.

At Bell Atlantic, economic data available from the government for political or statistical areas are disaggregated into user defined telephone service areas through the use of population weights derived from the centroid point assignment process described above, or the FIPS code process. Since economic data are available from the government for the whole counties contained in the LATAs, disaggregation only occurs in the case of split counties. The census tract components of counties are assigned to their respective LATAs using the procedures outlined in 1. and 2. above.

The census profiles developed through the use of the geographic linking systems provide the basis for developing time series data and forecasts of population through iterative proportional fitting schemes, when linked to historic and forecast information for the aggregates of counties which correspond to the LATAs. Economic data, in turn can be derived through the use of the time series and forecast versions of the population weights.

Thanks to the geographic linking processes developed jointly by the government and private sector firms with software capabilities, Bell Atlantic is able to address problems in the major corporate functional areas outlined earlier in this paper, utilizing public data as they relate to our odd geographic areas. □

Attrition and the National Longitudinal Surveys of Labor Market Experience: Avoidance, Control and Correction

by Dr. Patricia Rhoton¹
Data Archivist
Center for Human Resource Research
The Ohio State University

Since 1966 the Center for Human Resource Research has been analysing the longitudinal surveys conducted by the Census Bureau for the Department of Labor. The main purpose of

these surveys is to study the labor force activity of different population groups. The original groups included men who were 45–59 years old in 1966, women who were 30–44 years old in 1967, men who were 14–24 years old in 1966, and women who were 14–24 years old in 1968. In 1979, a new survey, conducted by the National Opinion Research Center (*editor's note:*) (NORC) in Chicago, was added for young men and women who were 14–21 in that year. Each of the five surveys is designed to collect information on all phases of the respondent's labor force activity and on other characteristics such as educational attainment, health, family composition, and financial status that are known to be related to such activity.

The original plan in 1965 was to interview the same respondents each year for a period of five years. Because of the usefulness of the data and the relatively small sample attrition, a decision was made at the end of the first five-year period to continue for another five years. The interview pattern was changed at that time from a face to-face yearly interview to a 2–2–1 pattern. Each respondent was contacted by phone every two years, then again in person one year after the second phone interview. This pattern was used again both during the third five-year extension obtained in 1976 and during the fourth five-year extension, obtained in December 1982. At the time of the most recent extension, a study looking specifically at attrition within the different cohorts was carried out.

Longitudinal studies in general have several advantages over the more frequent cross-sectional studies. While longitudinal studies are very expensive, the data are collected in great detail over time, with respondents reporting events and attitudes as they occur rather than retrospectively. Collecting the data in this way also enables the researcher to go beyond issues of correlations to address the more urgent issues of causality. The main advantage of a longitudinal survey, following the

¹This report was prepared under a contract with the Employment and Training Administration, U.S. Department of Labor, under the authority of the Comprehensive Employment and Training Act. Researchers undertaking such projects under government sponsorship are encouraged to express their own judgments. Interpretations or viewpoints stated in this report do not necessarily represent the official position or policy of the U.S. Department of Labor.

same set of respondents year after year, creates two major problems, however. The first is the difficulty of relocating respondents for subsequent interviews, and the second is maintaining respondent cooperation over repeated interviews.

Attrition in the NLS

Table 1² shows the numbers and percentages of respondents for all interviews up to and including the 1983 questionnaire. The base year row shows only those respondents who were interviewed that first year. Between the original screening and the first interview, some of the eligible respondents were lost: 9.0 percent of the Older Men, 5.5 percent of the Older Women, 8.3 percent of the Young Men, 5.8 percent of the Young Women, and 11.5 percent of the New Youth.

While Table 1 shows the distribution of interviews between and among the five cohorts, Tables 2–5 show interview/noninterview status of the four older cohorts by reason for noninterview. While there are differences between the cohorts in the distribution of reason for noninterview, within each cohort the distribution of reason remains consistent across the years. The method of interview, whether face-to-face or by telephone, does not seem to affect the attrition rate. Some of the losses in the sample are unavoidable. In the Survey of Mature Men (Table 2), for example, an increasing percentage of sample losses are due to respondent deaths. The Mature Women's survey (Table 3) has the second highest retention rate among the four older cohorts. This high rate is probably due to the fact that this group is very stable and has low geographic

mobility.

The Young Men's cohort has the lowest rate of retention and has been the test case for new attempts to stop the gradual decline in sample size. A variety of factors account for the difficulty in locating these respondents: completion of school, acquisition of new jobs, formation of families, and movement in and out of the military services. The higher rates of attrition in the earlier years were attributed to influx into the military since the sample was drawn, and initial interviewing done during the Vietnam War. However, rates remained high even as these respondents returned from the military.

The Young Women's cohort, which is similar to the Young Men's with respect to completion of school, acquisition of new jobs, and formation of families, posed the added challenge of name changes accompanying changes in marital status, yet the overall response rate has remained high.

The New Youth cohort has benefited greatly from the lessons taught by experience with the four older cohorts. In 1983, the response rate for this group was 96.3 percent. A comparison between this cohort and the first five years of the Young Women's cohort, which had the best retention rate of the older cohorts, shows that different procedures and techniques can substantially decrease attrition.

Not only does NORC have a higher overall interview rate, but also the organization seems to be better at retrieving respondents. In 1982, 96.0 percent of the original 1979 sample were interviewed. Some of these had not been interviewed in previous years: 2.2 percent in 1980, 1.1 percent in 1981, and 0.5 percent in 1980 or 1981. Only 165 respondents (one percent) of the original sample had had only one interview after four rounds of the survey. In 1983, the number of respondents who had had only one interview dropped to 115. Over eleven thousand (90.7) respondents had been

²Editor's note: Tables are gathered together at end of article

interviewed every year, and 5.5 percent had completed four out of the five interviews.

The Impact of Attrition on Representativeness

The gradual decline in sample size over time becomes very important if it results in a biased sample. While each cohort was checked at the end of the first five-year series of interviews, and smaller checks were made in the context of reports on occupational distribution, educational attainment, age distributions, and marital status with published national data, no one looked at all the cohorts systematically until 1982. At this point the issue of representativeness had to be addressed as part of the proposal to extend the cohorts for another five years.

Such a study could essentially be carried out in either of two ways. First, the remaining sample could be compared with some outside group, such as the decennial Census or the Current Population Survey. Comparison with an outside sample was difficult given time constraints and the fact that Census data were not yet ready for release. While the CPS data were available, differences between the CPS and each of the four older cohorts had already been documented in the first year. The second alternative was to compare the characteristics of the respondents who were left after ten years with the characteristics of all respondents interviewed in the initial year to see how much difference, if any, there actually was. Each cohort was checked for differences in the age distributions, educational attainment levels, employment status, industry and occupation distributions, marital status, SMSA status, annual income distribution, and wage and salary distribution. The Young Men and Young Women were also checked for differences in enrollment status.

A separate evaluation was done by race for each of the four cohorts. Table 6 is an example of the type of table constructed for each group. The ten-year sample was weighted using two methods: the entry level weight and a ten-year weight, which includes successive adjustments for each year's noninterviews. For all cohorts except the Young Men, the relevant comparison was between the entry year weighted figures and the ten-year sample using the ten-year weight. In the Young Men's cohort, the 1966 sample using the 1966 weight was compared to the 1976 sample using the 1966 weight because the 1976 weight had been adjusted to include individuals formerly in the military. Since young men already in the military had been deliberately excluded from the Young Men's sample, using the 1976 weight could have created apparent differences where none existed. For this group alone, it was more appropriate to use the 1966 weight.

Table 7 summarizes the distribution of differences by cohort and shows that for most characteristics the difference between the two samples was less than two percentage points. After the differences were identified, statistical tests of significance were computed for each of the comparisons. Table 8 shows the number of statistically significant differences at various levels for each cohort by race. While the number of differences was higher than would be expected by chance, several were based upon small sample cases in the initial year and characteristics with only two values. In the latter cases, a statistically significant result in one category means the other category will also be statistically different [sic].

After reviewing the entire set of tables, it was clear that noninterviews had not seriously distorted the representativeness of the sample. Given this finding, and the ability to apply weights to eliminate any potential bias, the decision was made to continue all four surveys for another five years.

It is unclear, however, how further erosion of the samples will affect representativeness. Concern with this issue, together with the higher noninterview rates that NORC was having with the New Youth sample, led to an evaluation of the rules that had been established in the original five-year period and an attempt to see if it was possible to retrieve some of the noninterview cases.

Retrieving Former Noninterview Cases

Since the Young Men's panel had lost the most respondents, it became the target for the first attempt at retrieval. Respondents from the 1975, 1976, 1978 and 1980 survey years, who normally would not have been included in the workload (i.e., attempted contacts) because of noninterview status in those years (refused, unable to contact, institutionalized, moved outside the U.S.) were sorted, and a sample of 279 selected.

Several changes were made in the procedures for contacting these special respondents. No restrictions were placed on the number of telephone calls, mileage, or time spent locating and retrieving these respondents. Each interviewing packet included the respondent's most recently completed interview and household record card, as well as the most recent questionnaire, and all record cards for any other household members participating in any of the other cohorts. In addition, an expanded list of methods of locating respondents was included. As a result of these additional steps, 104 (37.3 percent) of these respondents were interviewed. These interviews have been flagged and will be checked as soon as the data tapes are available from the Census Bureau to determine if they differ in any way from the rest of the respondents. If these respondents remain in the sample for the next round of

interviews in the latter part of 1983, a concerted effort may be made to use these procedures during the regular interviews and in similar attempts to retrieve noninterviews in the other three cohorts.

Differences Between Census and NORC

One of the major differences between Census and NORC is the amount of location information obtained from the respondent. NORC obtains more information, and request information on other individuals with specific relationships to the respondent, depending upon the respondent's circumstances. The interviewer begins by asking the name, relationship, address, and phone number of the person most likely to know where the respondent is. If the respondent is living in a dormitory, fraternity, sorority, hospital or other temporary situation, the interviewer is instructed to obtain the name and relationship of a householder at a permanent home address. If the respondent is married and living apart from a spouse, the spouse's address and telephone number are requested. If the respondent is not living with a parent and has not provided a parent's name, this information is obtained, including whether or not the parents live together. The name of another relative with whom the respondent is in contact, and the names of friends and places to which the respondent goes when not spending spare time at home, are also obtained. Respondents are also asked for nicknames, maiden names if they are married women, and whether or not they expect to move in the next 12 months.

This extensive list gives the NORC interviewer a real advantage when contacting someone on the list, since the ability to mention the respondent's parents, relatives, friends, hangouts or nicknames demonstrates that the interviewer

knows the respondent to some degree and may make the reference more willing to give out information about the respondent. Another major advantage that the NORC interviewer has over the Census interviewer is the existence of a centralized locating shop in Chicago. The person working at the locating shop has access to all previous questionnaires, original copies of locator documents and information about the respondent's brothers and sisters. Working with this additional data, the respondent can usually be located by phone and reassigned to the same or another interviewer. The Census interviewer starts out with less information with which to locate the respondent. S/he has a questionnaire with a label indicating the respondent's name and most recent home address. In addition, there is a household record card for each respondent which contains the telephone numbers, all addresses at which the respondent has lived since the survey began, the names of all persons who have lived with the respondent, and the names, addresses and telephone numbers of only two persons who will always know where s/he can be reached.

Besides the more extensive locating supplement that NORC builds in the interview, several other differences appear. Each respondent in the New Youth cohort is paid \$10.00 for a completed interview, since many researchers believe that even a small amount of money helps in obtaining cooperation, especially among younger respondents. The New Youth respondents also had an opportunity to take a series of tests for the Department of Defence, which needed to evaluate tests given to individuals in the military. For these tests, which take several hours, the respondents were paid \$50.00. When the four older cohorts were first interviewed, paying respondents was not as well accepted. Now there are fears that starting this procedure with the older cohorts would cause concern on the part of respondents.

Another procedural difference is that New Youth cohort respondents are told up front that

they will be interviewed each year for the next several years and are therefore aware that they will be contacted about the same time each year. The Census interviewers are told only that they may be conducting additional surveys, and should not tell the respondents that this is the last time s/he will be interviewed. The lack of an answer to give the respondent, in addition to the 2-2-1 pattern, probably leaves the respondent without a sense of when or if s/he will be contacted again. While this ambiguity may not have an impact on their cooperation in the survey, the NORC approach leaves the respondent with a greater feeling of certainty about the interviewing schedule.

Revising the Rule for Dropping Respondents

After the first year, respondents in the four older cohorts who refused to participate or had died, were dropped from the Census sample. Those who were not reinterviewed for any reason for two consecutive years were also dropped. The only exception was made in the Young Men's sample with those respondents who were in the Armed Forces. Since the sample was to represent the national civilian, non-institutionalized population, young men were not interviewed while they were in the Armed Forces but were retained in the sample and reinterviewed in the first interview after they had left the services. However, NORC's success in retrieving respondents even after they had refused and the success of the Young Men retrieval effort resulted in a change in these rules. Currently, no respondent is dropped except those who have died. NORC goes back each year and attempts to interview all living respondents.

Maintaining Respondent Cooperation

While both Census and NORC send out advance letters about the entire survey, stressing the importance of the respondent's cooperation, NORC also sends out a newsletter that tells respondents in a very "chatty" format about some general results of the previous survey. The Census Bureau had a short, formal fact sheet that went out with the cover letter, but interviewers reported that respondents did not feel it was very useful. In the 1982 Young Women's survey, a more extensive description of the surveys and a list of the research results from the survey were sent to any respondent who filled in and returned a postcard requesting additional information. Over one-third of the respondents interviewed in that wave mailed in the postcard. A variable will be created identifying these respondents and if distribution of the handbook increases the response rate in the next round, the handbook will be offered to the respondents in all three cohorts.

Conclusions

The New Youth survey has, at this time, a considerably better response rate than any of the four older cohorts. Much of its success can be traced to the solution of problems that developed over time in the four older cohorts. While the necessity of maintaining the same measures over time prevented change in the handling of the four older cohorts, these problems were corrected in the first wave of the New Youth cohort. Questions that the respondents or the interviewer had difficulty with in the four older cohorts were altered so that there was no confusion from the very beginning. Perhaps most importantly, given the highly mobile nature of the younger age group, much more detail was obtained on individuals who would always know where the respondent was. In addition, more information about the survey was given to the respondent before, during and after each interview. All these factors combined have resulted in a response rate that is very good for any survey and exceptional for a longitudinal survey in its fifth year. □

Table 1
Table 1 National Longitudinal Surveys Retention Rates

Year	Apr 35-59 in 1966		Apr 50-64 in 1967		Apr 14-24 in 1968		Apr 14-24 in 1968		Apr 14-24 in 1968		Apr 14-21 in 1979		Apr 14-21 in 1979	
	Total interviewed	Retention rate	Total interviewed	Retention rate	Total interviewed	Retention rate	Total interviewed	Retention rate	Total interviewed	Retention rate	Total interviewed	Retention rate	Total interviewed	Retention rate
1966	5,034	100.0			5,225	100.0								
1967	4,751	94.4	5,083	100.0	4,790	91.7								
1968	4,661	92.6	4,910	96.6	4,318	82.6	5,159	100.0						
1969	4,388	87.2	4,712	92.7	4,033	77.2	4,530	95.6						
1970					3,993	76.4	4,766	92.4						
1971	4,182	83.1	4,575	90.0	3,987	76.3	4,714	91.4						
1972			4,471	88.0			4,625	89.6						
1973	3,951	78.5			4,014	76.8	4,424	85.8						
1974			4,322	85.0										
1975	3,732	74.1			3,977	76.1	4,243	82.2						
1976	3,487	69.3	4,172	82.1	3,696	70.7								
1977			3,966	78.0			4,071	78.9						
1978	3,219	63.9			3,538	67.7	3,923	76.1						
1979			3,812	75.0			6,298	100.0	6,288	100.0				
1980	3,001	59.6			3,438	65.8	3,801	73.7	6,087	95.1	6,054	96.3		
1981	2,834	56.2	3,677	72.3	3,398	64.9			6,125	95.7	6,070	96.5		
1982			3,542	69.7			3,650	70.8	6,083	95.1	6,040	96.1		
1983	2,634	52.3					3,545	68.7	6,143	96.0	6,078	96.7		

Table 2

Table 2 Interview/Noninterview Status for Survey of Mature Men

Year	Total inter-views	Noninterview				Interview				Armed forces #	Institutionalized #	Other #	Noninterview 2 consecutive years dropped	Retention rate %
		Total #	Decensored #	Refused #	Unable to contact #	Temporarily absent #	Temporarily absent %	Institutionalized %	Armed forces %					
1966	5,034	-	-	-	-	-	-	-	-	-	-	-	-	-
1967	4,751	283	5.6	60	21.2	107	37.8	93	32.9	8	2.8	-	-	94.4
1968	4,661	206	4.2	75	36.4	50	24.3	66	32.0	3	1.5	1	0.5	92.6
1969	4,388	330	7.0	102	30.9	92	27.9	88	26.7	17	5.1	13	3.9	87.2
1970	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1971	4,182	323	2.2	169	52.3	56	17.3	56	17.3	15	4.6	17	5.3	19
1972	-	-	-	-	-	-	-	-	-	-	-	-	-	83.1
1973	3,951	303	2.1	175	57.7	42	13.9	62	20.5	3	1.0	14	4.6	26
1974	-	-	-	-	-	-	-	-	-	-	-	-	-	78.5
1975	3,732	281	7.0	174	61.9	43	15.3	37	13.2	3	1.1	13	4.6	24
1976	3,487	284	7.8	100	34.0	128	43.5	22	7.5	18	6.1	14	4.8	15
1977	-	-	-	-	-	-	-	-	-	-	-	-	-	69.3
1978	3,219	316	6.9	163	51.6	73	23.1	38	12.0	12	3.8	26	8.2	14
1979	-	-	-	-	-	-	-	-	-	-	-	-	-	63.9
1980	3,001	289	8.8	202	69.9	35	12.1	15	5.2	2	0.7	29	10.0	13
1981	2,834	202	6.6	99	49.0	58	29.2	11	5.4	5	2.5	26	12.9	56.2

Percent of previous year number of interviews.

¹The retention rate in the percent of base year re-

3. Noninterviews which are deleted from the sample.

The retention rate in the percent of base year respondents who were interviewed.

Table 3

Table 3 Interview/Noninterview Status for Survey of Mature Women

Year	Total inter-views	Noninterview																	
		Total #	% ¹	Decensed #	% ²	Refused #	% ³	Unable to contact #	% ⁴	Temporarily absent #	% ⁵	Institutionalized #	% ⁶	Armed forces #	% ⁷	Other #	% ⁸	Noninterview 2 consecutive years dropped	Retention rate ⁹
1966	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1967	5,083	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1968	4,910	173	3.4	22	12.7	76	43.9	62	35.8	-	-	1	0.6	-	-	12	6.9	-	96.6
1969	4,712	273	5.5	13	4.8	134	49.1	87	31.9	16	5.9	7	2.6	-	-	16	5.9	-	92.7
1970	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1971	4,575	234	4.9	25	10.7	82	35.0	84	35.9	12	5.1	6	2.6	-	-	25	10.7	24	90.0
1972	4,471	199	4.3	12	6.0	97	48.7	66	33.2	8	4.0	2	1.0	-	-	14	7.0	46	88.0
1973	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1974	4,322	196	4.3	29	14.8	90	45.9	51	26.0	8	3.1	5	2.6	-	-	15	7.7	39	85.0
1975	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1976	4,172	212	4.8	30	14.2	101	47.6	48	22.6	5	2.4	7	3.3	-	-	21	9.9	14	82.1
1977	3,965	267	6.3	9	3.4	181	67.8	41	15.3	10	3.7	6	2.2	-	-	20	7.5	19	78.0
1978	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1979	3,812	201	5.0	36	17.9	106	52.7	35	17.4	4	2.0	4	2.0	-	-	16	8.0	22	74.9
1980	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1981	3,660	170	4.4	37	21.8	91	53.5	26	15.3	4	2.4	6	3.5	1	0.5	5	2.9	-	72.4

¹Percent of previous year number of interviews.
²Percent of total noninterviews.
³Noninterviews which are deleted from the sample.
⁴The retention rate in the percent of base year respondents who were interviewed.

Table 4

Table 4 Interview/Noninterview Status for Survey of Young Men

Year	Total inter-views	Noninterview									
		Total #	% ¹	Deceased #	Refused #	Unable to contact #	Temporarily absent #	Institutionalized #	Armed forces #	Other #	Noninterview 2 years dropped #
1966	5,225	-	-	-	-	-	-	-	-	-	-
1967	4,790	435	8.3	14	3.2	83	19.0	4	0.9	-	-
1968	4,318	827	16.1	10	1.2	70	8.5	143	17.3	6	0.7
1969	4,033	994	19.8	11	1.1	54	5.4	181	18.2	10	1.0
1970	3,993	904	18.5	13	1.4	37	4.1	154	17.0	17	1.9
1971	3,982	801	16.7	11	1.4	54	6.7	160	20.0	13	1.6
1972	-	-	-	-	-	-	-	-	-	-	-
1973	4,014	655	14.0	17	2.6	69	10.5	214	37.3	6	0.9
1974	-	-	-	-	-	-	-	-	-	-	-
1975	3,977	552	12.2	13	2.4	88	15.9	212	38.4	1	0.2
1976	3,696	541	12.8	7	1.3	177	32.2	152	28.1	9	1.7
1977	-	-	-	-	-	-	-	-	-	-	-
1978	3,538	472	11.8	21	4.4	114	24.1	164	34.7	8	1.7
1979	-	-	-	-	-	-	-	-	-	-	-
1980	3,438	408	11.9	15	3.7	72	17.6	165	40.4	5	1.2
1981	3,398	391	7.5	7	1.8	107	27.4	126	32.2	8	2.0

¹Percent of previous year number of interviews.²Percent of total noninterviews.³Noninterviews which are deleted from the sample.⁴The retention rate in the percent of base year respondents who were interviewed.

Table 5

Year	Total inter-views	Noninterview					Total					Noninterview 2 consecutive years dropped	Retention rate ^a		
		Total #	Percent of %2	Refused #	Unable to contact #	Ten- per- cent #	Institu- tionalized %2 #	Armed forces %2 #	Other #						
1966	-	-	-	-	-	-	-	-	-	-	-	-	-		
1967	-	-	-	-	-	-	-	-	-	-	-	-	-		
1968	5,159	-	-	-	-	-	-	-	-	-	-	-	-		
1969	4,930	229	4.4	2	0.9	98	42.8	112	48.9	3	1.3	9	3.9	-	95.6
1970	4,766	293	5.8	6	2.0	74	25.3	136	46.4	21	7.2	7	2.4	-	92.4
1971	4,714	217	4.4	4	1.8	54	24.9	95	43.8	12	5.5	8	3.7	-	91.4
1972	4,625	203	4.2	6	3.0	74	36.5	93	45.8	5	2.5	5	2.5	-	89.7
1973	4,424	285	6.1	4	1.4	116	40.7	108	37.9	4	1.4	4	1.4	-	85.5
1974	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1975	4,243	307	6.7	5	1.6	114	37.1	132	43.0	5	1.6	2	0.7	-	82.2
1976	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1977	4,108	274	6.3	3	1.1	113	41.2	114	41.6	5	1.8	2	0.7	-	79.6
1978	3,902	319	7.5	3	0.9	180	56.4	86	26.9	14	4.4	6	1.9	-	75.6
1979	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1980	3,801	190	4.8	6	3.1	88	46.3	63	33.1	1	0.5	3	1.6	-	73.7

³Noninterviews which are deleted from the sample.

⁴The retention rate in the percent of base year respondents who were interviewed.

Percent of previous year number of interviews.

¹Percent of previous year number
²Percent of total noninterviews.

Table 6

Table 6 Selected Characteristics in 1966 of Original Sample and Sample Interviewed in 1976

Mature Men - Whites Only															
Characteristics in 1966	Number of respondents in 1966	# potentially eligible for interview in 1976 ¹		Number of respondents in 1976	Total retention rate	1966 sample				1976 sample					
						Unweighted		Weighted using 1966 weight ¹		Unweighted		Weighted using 1966 weight ¹			
						#	%	#	%	#	%	#	%		
		#	%			#	%	(000)	%	#	%	(000)	%	(000)	
Age															
45-49	1329	1202	39.3	983	74.0	1329	36.9	4996	36.5	983	39.0	3691	38.7	4482	38.8
50-54	1230	1043	34.1	868	70.6	1230	34.2	4638	33.9	868	34.4	3239	34.0	3917	33.9
55-59	1041	811	26.5	670	64.4	1041	28.9	4041	29.6	670	26.6	2610	27.4	3152	27.3
Educational attainment															
Less than 12 yrs.	2038	1679	55.3	1391	68.2	2038	57.0	7645	56.2	1391	55.3	5198	54.6	6294	54.6
12 years	885	778	25.6	652	73.7	885	24.7	3385	24.9	652	25.9	2486	26.1	3002	26.1
More than 12 yrs.	655	583	19.2	472	72.1	655	18.3	2561	18.8	472	18.8	1834	19.3	2227	19.3
Employment status															
Employed	3348	2897	94.9	2395	82.7	3348	93.0	12709	93.0	2395	95.0	9056	94.9	10561	94.9
Unemployed	46	38	1.2	27	71.0	46	1.3	172	1.3	27	1.1	106	1.1	130	1.1
Out of labor force	206	121	4.0	99	81.8	206	5.7	789	5.8	99	3.8	377	4.0	457	4.0
Industry ²															
Agriculture	335	293	10.1	264	90.1	335	10.0	1171	9.2	264	11.0	920	10.2	1100	10.0
Mining	30	25	.9	23	92.0	30	0.9	118	0.9	23	1.0	91	1.0	110	1.0
Construction	351	292	10.1	237	81.2	351	10.5	1335	10.5	237	9.9	900	10.0	1094	10.0
Manufacturing	1000	866	29.9	702	81.1	1000	29.9	3805	30.0	702	29.3	2660	29.4	3220	29.4
Transportation	315	263	9.1	225	85.6	315	9.4	1213	9.6	225	9.4	858	9.5	1036	9.5
Trade	523	457	15.8	375	82.0	523	15.6	2008	15.8	375	15.7	1441	15.9	1746	15.9
Finance	129	111	3.8	89	80.2	129	3.9	509	4.0	89	3.7	351	3.9	425	3.9
Services	432	374	12.9	301	80.5	432	12.9	1644	13.0	301	12.6	1142	12.6	1388	12.7
Public admin.	230	213	7.4	177	83.1	230	6.9	896	7.1	177	7.4	685	7.8	832	7.6
Marital status															
Married	3235	2775	91.0	2298	82.8	3235	90.1	12278	90.0	2298	91.3	8695	91.3	10526	91.3
Previously married	190	141	4.6	119	84.4	190	5.3	732	5.4	119	4.7	453	4.8	551	4.8
Never married	166	134	4.4	101	75.4	166	4.6	624	4.6	101	4.0	379	4.0	457	4.0
Occupation ²															
Professional	359	316	10.9	254	80.4	359	10.8	1406	11.1	254	10.8	999	11.1	1218	11.1
Managerial	582	508	17.6	424	83.5	582	17.4	2245	17.7	424	17.7	1623	18.0	1969	18.0
Clerical	173	147	5.1	118	80.3	173	5.2	667	5.3	118	4.9	448	5.0	542	5.0
Sales	176	154	5.3	125	81.2	176	5.3	688	5.4	125	5.2	491	5.4	597	5.5
Crafts	878	720	24.9	595	82.6	878	24.8	3153	24.9	595	24.9	2249	24.9	2722	24.9
Operatives	572	497	17.2	400	80.5	572	17.1	2198	17.3	400	16.7	1544	17.1	1867	17.1
Household	180	149	5.2	123	82.6	180	5.4	663	5.2	123	5.1	446	4.9	541	5.0
Services	255	228	7.9	206	90.4	255	7.6	889	7.0	206	8.6	718	7.9	857	7.8
Farmers	25	21	.7	19	92.0	25	0.7	95	0.7	19	0.7	71	0.7	86	0.7
Farm laborers	60	50	1.7	46	92.0	60	1.8	212	1.7	46	1.9	159	1.8	191	1.8
Laborers	154	119	4.1	98	82.4	154	4.6	556	4.4	98	4.1	356	3.9	429	3.9
SSEA status															
In SSEA	2487	2107	69.0	1716	81.4	2487	69.1	9506	69.5	1716	68.1	6550	68.7	7945	68.8
Out of SSEA	1112	948	31.0	804	84.8	1112	30.9	4163	30.5	804	31.9	2988	31.3	3601	31.2
Annual income															
Equal to zero	5	5	.2	5	100.0	5	0.2	19	0.2	5	0.3	19	0.3	24	0.3
1-2,999	249	185	7.9	157	84.9	249	9.1	916	8.7	157	8.0	569	7.6	684	7.5
3,000-9,999	1418	1177	50.5	999	84.9	1418	51.5	5367	51.0	999	50.6	3764	50.2	4555	50.1
10,000-14,999	712	631	27.1	534	84.6	712	25.9	2765	26.3	534	27.1	2069	27.6	2510	27.6
15,000-19,999	711	192	8.2	153	79.7	711	7.7	845	8.1	153	7.8	604	8.1	735	8.1
+ 20,000	156	141	6.0	125	88.6	156	5.7	596	5.7	125	6.3	477	6.4	577	6.4
Wages and salary															
Equal to zero	777	634	27.5	541	71.1	777	23.5	2876	23.0	541	22.3	2009	21.9	2420	21.8
1-2,999	243	189	6.7	172	91.0	243	7.2	899	7.2	172	7.1	624	6.8	755	6.8
3,000-9,999	1696	1468	52.1	1262	86.0	1696	51.3	6451	51.5	1262	52.1	4790	52.2	5801	52.2
10,000-14,999	447	397	14.1	343	86.4	447	13.5	1732	13.8	343	14.2	1326	14.5	1614	14.5
15,000-19,999	81	77	2.7	58	75.3	81	2.5	333	2.7	58	2.4	238	2.6	290	2.6
+ 20,000	60	54	1.9	47	87.0	60	1.8	235	1.9	47	1.9	185	2.0	226	2.0

¹Excludes death, military and out of country.²Those employed survey week.

Table 7

Table 7 Number and Percentage of Differences by Panel

Panel	Absolute differences (%)			Total
	0-2	2-3	3+	
Mature men				
Black	34 (73.9)	8 (17.4)	4 (8.7)	46 (100.0)
White	43 (95.6)	2 (4.4)	0	45 (100.0)
Mature women				
Black	42 (93.3)	3 (6.7)	0	45 (100.0)
White	45 (100.0)	0	0	45 (100.0)
Young men				
Black	30 (73.2)	5 (12.2)	6 (14.6)	41 (100.0)
White	43 (97.7)	1 (2.3)	0	44 (100.0)
Young women				
Black	33 (82.5)	6 (15.0)	1 (2.5)	40 (100.0)
White	40 (95.2)	2 (4.8)	0	42 (100.0)

Table 8

Table 8 Number and Percentage of Statistically Significant Differences by Panel

Panel	Level of significance		
	1%	2%	3%
Mature men			
Black	4 (9.1)	7 (15.9)	12 (27.3)
White	4 (9.1)	7 (15.9)	14 (31.8)
Mature women			
Black	2 (4.5)	2 (4.5)	3 (6.8)
White	1 (2.3)	4 (9.3)	5 (11.6)
Young Men			
Black	1 (2.6)	4 (10.3)	6 (15.4)
White	2 (4.7)	4 (9.3)	6 (14.0)
Young women			
Black	1 (2.6)	3 (7.9)	4 (10.5)
White	1 (2.6)	2 (5.1)	2 (5.1)

CAST: Centre for Applications Software and Technology

by A. Stacey¹
CAST
Edinburg, UK

CAST/EUL Data Library Services

The Data Library at the University of Edinburgh offers the research community on-line access to facilities for the analysis of census, time-series, sample survey, and related data. These data are held on the Edinburgh Regional Computing Centre's network of ICL 2900 and VAX mainframe computers, in an environment that is well suited to (remote) multi-access interactive computing, and is rich in software for statistical analysis and graphic display. The 'official statistics' data holdings of the Data Library include:

A very full range of (small) area statistics relating to the 1971 and 1981 Population Censuses for Scotland, England and Wales;

A series of Parish summary and grid

square data from the Annual Agricultural Census, for England and Wales;

The CSO Macroeconomic Databank;

The Scottish Input/Output Tables for 1979;

The OS Gazetteer;

The Postcode Directory for Scotland;

A full range of digitised boundary files.

Information about the holdings of the Data Library are held in an on-line view system, DATALIB. DATALIB and the holdings themselves are accessible on-line by users from the university communities of Edinburgh, Glasgow and Strathclyde, and by a range of users across the PSS/JANET computer networks, or across the telephone network (using an acoustic coupler/modem). In providing an on-line data library service, the Data Library relies upon the expertise, facilities and staff of the Edinburgh Regional Computing Centre (ERCC), the Edinburgh University Library (EUL) and the Centre for Applications Software and Technology (CAST). It also draws heavily upon the past experience of what was known as the Program Library Unit (PLU).

Access to the BUSH mainframe on the ERCC Network (EDNET) is as follows:

PSS : 2342 313 54354 (and then CALL BUSH)

JANET: 0000 0700 1004 04 (and then 1500 0003 or CALL BUSH)

Datel: (UK) 031-667-1071

[In reply to the prompt 'User' enter

¹Paper prepared for presentation at the IFDO/IASSIST Conference, Workshop on Providing Local Data Services, Amsterdam, May 20 - 23, 1985.

DATALIB]

The Data Library and PLU

The Data Library formally came into being in the mid-1970's in order to provide computing facilities and access to small area statistics from the 1971 Population Census (Scotland). The machine-readable data from this Census were purchased for Edinburgh University's academic research and teaching community by the University Library. The data were managed and were made accessible on the ERCC's network of mainframe computers through programs written by the staff of (what was then called) the Program Library Unit (PLU). PLU was a specialist unit within Edinburgh University with a national role for the conversion and maintenance of statistical packages for ICL computers in universities. It had also gained considerable experience in writing census-access software and had familiarity with computer-aided mapping from work on the annual Agricultural Census (England and Wales) which it had carried out in association with the Department of Geography. (The mapping programs, CAMAP and GIMMS are in widespread use today.) Four programs were written for accessing the 1971 Population Census, together with DATAPAC, a user-friendly interface designed with the novice computer user ('whose expertise lies outwith [sic] computing') in mind.

The PLU played a significant role, in the U.K., in the dissemination and promotion of secondary analysis of (small) area statistics from the 1981 Population Censuses. First, the application software that was commissioned by LAMSAC (Local Authorities Management Services Advisory Council) for the retrieval and

manipulation of machine-readable tabular output from the Census was designed and written (under subcontract to Durham University) by staff of the PLU. The SASPAC project, as it was called, won the British Computer Society's Social Benefit Award, and is widely used by local government officers and academic research staff.

Second, PLU played an active part in the two consortia which, on behalf of the academic community, purchased the 1981 Population Census statistics from the offices of the two Registrars-General. In particular, PLU formed the consortium to purchase the data from the Scottish Census, and arranged for these to be deposited with the ESRC Data Archive for general distribution. Later, as part of the Inter-University Software Committee (IUSC)'s working party on census data, the Data Library at Edinburgh became one of the six regional and national computer centres to act as a census data library for academic research purposes. The General Register Office (Scotland) also granted a Census Agency Agreement to the University of Edinburgh to enable the Data Library to provide services to commercial users, including academics conducting contract research and policy analysts in central and local government. PLU has also collaborated in academic research projects. These have included the computerisation of a 2% sample of the enumerators books from the 1851 Population Census, and provision for public access to these data through, for example, the distribution of floppy disks for viewing on the BBC microcomputer.

CAST, EUL and ERCC

The Centre for Applications Software and Technology (CAST) was established in 1983 as an outcome of a review of the Program Library Unit, occasioned by the retirement of its Director. CAST now has a staff of about 40: in addition to what are now referred to as the Data Library Service and the Program Library Service, CAST also has expertise in database design and management, numerical algorithms, statistical computing, graphics, application software evaluation, conversion and development (for mainframe and micro computers) and survey methodology. (CAST continues to provide some software services under the name PLU.) The Data Library can therefore call upon specialists for advice or project work on a range of relevant activities. In particular, the provision of access software with attractive user interfacing (which has for example been written for the CSO Macroeconomic Databank, the Agricultural Census and the OS Gazetteer) is central to the development of the Data Library, which is why the service is housed at CAST.

The Library of the University of Edinburgh (EUL) is one of the major university libraries in the U.K. It has about 200 full-time and part-time staff, and a large reference collection including official statistics and maps. It is currently undertaking the on-line cataloguing of its holdings, and makes extensive use of on-line bibliographic search facilities (including DIALOG, ERIC, etc.). It has begun to grasp the nettle of cataloguing machine-readable data files, and many of the computer terminals installed in the library for use by its 'users' are terminals also connected to the local computing network. Both the official statistics and the map collections provide necessary and complementary services to the Data Library. For example, researchers wishing to use the small area statistics from the Census may consult the maps in the map reference section

in order to discover where enumeration district boundaries lie, for example, and consult the published tables from the Census in the statistical reference section prior to conducting an analysis of the machine-readable small area statistics, and perhaps producing a high quality schematic map of their own.

The Edinburgh Regional Computing Centre (ERCC) was founded in 1966, and has maintained on its network of mainframe computers an operating system called the Edinburgh Multi-Access System (EMAS), specifically designed to provide an interactive computing environment for a scattered population of users. This was considerably in advance of the 'star' network systems which were later developed at national and regional sites. The ERCC therefore has considerable experience in both interactive and batch computing, the provision of on-line help information and experience in communications between different mainframes. The EMAS operating system can also claim to be a particularly 'friendly' operating system, when compared to the alternatives available, especially for the first time user from another computing site.

General Institutional and Technological Environment

The University of Edinburgh, with over 1500 teaching and research staff and over 10,000 students, has a widely dispersed campus. It is partly for this reason that Edinburgh has a telecommunications network that is arguably one of the most advanced among the universities in the U.K. The Edinburgh Multi-Access System (EMAS) has provided a powerful but 'friendly' interactive and batch computing environment since 1972. EMAS resides, as an operating system, on the ICL 2988 and dual ICL 2976

mainframe computers, the latter being connected to two ICL Distributed Array Processors (DAP), and has just been introduced on a newly installed Amdahl mainframe. The EDNET network is a multi-mode packet switching network linking EMAS with VMS and UNIX operating systems on a range of mini-computers, and providing access facilities for the staff and students of Strathclyde and Glasgow universities. The network also offers access to a central (disk) filestore, printing and graphical facilities, an electronic mail service, and a (regional) gateway to the British Telecom Package Switching System (PSS) and the U.K. academic network (JANET). This is shown in the figure below. ERCC is now experimenting with integrated speech and data facilities on its network, evaluating the merits of early access to the full ISDN (Integrated Service Digital Networking) being piloted by BT in London, Manchester and Birmingham.

Current Staff

The staff of the Data Library Service are drawn from the University Library (EUL) as well as from CAST. Staff with major responsibilities, and who regularly spend part of their time in this area, include a manager, a senior computing officer, an administrative/computing assistant, a reference librarian and three computing officers, the latter with specialist responsibilities for programming, computer-assisted cartography and spatial analysis. These staff have skills in database management, statistics, survey analysis and design, cartography and have experience working in the social sciences, government, and the physical sciences.

Work-in-progress

The Data Library at Edinburgh is currently undergoing a reorganisation in order to provide an expanded range of services. This is partly in preparation for the launch of a Scottish Data Centre, but is also partly in response to a wish to foster data library development more generally within the UK. The cataloguing of machine-readable data files has been identified as a priority area, along with the provision of an on-line union catalogue of data collections, indicating the existence and location of data. Some preliminary discussions have been held with Sue Dodd and with the staff of the ESRC Data Archive who have made considerable progress in the compilation of MARC-compatible study descriptions of the Archive's data collections. We also recognise that the development of data libraries depends crucially upon the existence of a national data clearing house, with secure long-term funding, and upon agreement on the data library/clearing house relationship. We therefore have an interest in promoting an organisation like the Data Archive at Essex.

One of the features of Edinburgh's Data Library is the provision of user-friendly interfaces to access software. We have also become conscious of the value of friendly interfaces to mapping packages, and are in the process of designing EASYMAP which will, for example, generate the commands to create a GIMMS map, using digitised boundary data from a machine-readable library.

Funding for these activities comes from two major sources. First, the University provides finance for a core of staff. Second, CAST is able to generate revenue from various external activities. These include, for example, data processing and access management for the Edinburgh District Council's 'Homeless Survey'; consulting to the Scottish Office, and analysis of

Australian Health Statistics

by Roger Jones ¹
Social Science Data Archives,
Australian National University,
Canberra, Australia.

Introduction

The impetus for this paper was a workshop held earlier this year to identify current problems associated with Australian health statistics and determine priorities for a national health data base. Generally, academic researchers in Australia have focussed on aetiology, that is the study of causes of diseases, and attention will be given to the data used to develop aetiological hypotheses and, ultimately, test these hypotheses. By identifying the major causes of avoidable death, disease or disability as well as possible with existing data, the gaps in the data will be identified and priorities for the collection of new data will be established.

¹Paper prepared for presentation at the IFDO/IASSIST Conference, Amsterdam, May 20 - 23, 1985.

Editors note: Unfortunately the accompanying tables were not submitted with this paper. We have therefore deleted all specific table references. With apologies to the author.

Data Requirements

Increased attention is being given to the need for improved occupational health surveillance in western societies. In Australia, concerns about potential health effects of exposure to the herbicide 2,4,5-T and its dioxin contaminant, Vietnam War service in general, proximity to atomic testing, past employment in uranium mines, and exposure to asbestos and lead are prominent among the issues that have focussed attention on occupational risks. The risks of cancer resulting from occupational exposures have been discussed widely, and the patterns of accidents, injuries and illnesses associated with industry and occupation examined.

Personal and lifestyle factors such as diet, cigarette smoking, alcohol consumption, stress and drug use are also now recognised as important 'risk factors' associated with the health status of the population. There is now evidence to support the view that the chronic, degenerative diseases such as heart disease and cancer which are the major killers of Australians substantially result from these socio-behavioural factors rather than simply being the diseases of old age.

The main causes of death and hospitalization in Australia, for people below the age of 45, are motor vehicle accidents and other accidents, poisoning and violence, including suicide. For older people the chronic diseases of ischaemic heart disease, stroke and other diseases of the circulatory system and various cancers predominate.

A number of factors are known to influence, or to be associated with these causes of death. The most important behavioural factors, in terms of the amount of resultant disease, are almost certainly smoking and alcohol consumption. The Commonwealth Department of Health suggested that there were 16,200

deaths in Australia associated with tobacco use in 1980, equivalent to 110 deaths per 100,000 population. The diseases with which smoking is associated are clearly indicated in the warning 'Smoking causes lung cancer, as well as heart and other lung disease' which the NH & MRC (editor's note:) (National Health and Medical Research Council) recommended to replace the less specific 'Smoking is a health hazard' on cigarette packets sold in Australia. Almost half the number of deaths attributed to alcohol are associated with road traffic accidents, from which disability and injury also result. Diet is also implicated in several of the major causes of alcohol-related deaths.

Descriptive Studies

Considerable difficulties arise when trying to establish causal links with diseases, such as the argument over smoking and cancer. Individual studies are rarely definitive and considerable time may elapse before sufficient evidence is available to assume a causal relationship. Such studies can also be expensive. Accordingly, less intensive descriptive studies are undertaken in the first instance, using readily available data sources from which crude measures may be derived in order to generate or provide an initial check of hypotheses. Fully developed analytical studies would only follow when the results of exploratory investigations are sufficiently suggestive and important to warrant the expense of more rigorous confirmatory research.

Information for general descriptive studies may be available from routine collections of health event data or from prevalence surveys of population samples.

Routine Health Collections

The routine health reporting systems in Australia are poorly developed in comparison to other developed countries, particularly at the national level. Under the federal/state system of government that operates in Australia, state governments are primarily responsible for the provision of hospital and health services within their own borders, and for the collection of health event data associated with these services. While all states administer similar collections, substantial variation exists in terms of coverage, scope and means of collection.

Mortality data play an important role as health status indicators. Data are compiled in each state by the Registrars of Births, Deaths and Marriages and causes of death coding is added by the Australian Bureau of Statistics (ABS) from medical certificates. At the present time, there is no national compilation, although health ministers have agreed to establish a National Death Index, subject to appropriate confidentiality legislation being enacted. To obtain Australia-wide mortality tapes at present, with personal identifiers removed, separate tapes for each state would have to be developed by the ABS and sent to the state registrars for release, subject to their approval, and re-aggregated by the researcher. Clearly this is a very cumbersome procedure. The main public source of data at present is the ABS publication 'Causes of Death'.

Hospital morbidity collections based on data collected on patients leaving hospitals are compiled by the health authorities in each state, providing limited information on the demographic background and illnesses of hospital in-patients. In all states, data are collected on separations of public patients from public hospitals, but private hospitals, psychiatric hospitals and nursing homes may be included in some states and not in others, so that

comparisons between states and aggregation across states cannot be accomplished from published results, although a uniform minimal data set could be achieved with access to the computerised records.

Records of primary contact between the public and health professionals such as private medical practitioners, community health centres, hospital clinics and casualty services are totally lacking at present, at least in any unified form. However, the introduction of the universal health insurance scheme, Medicare, in February 1984 has created an opportunity to obtain more information in this area. Medicare gives automatic entitlement to a subsidy of 85% of the schedule fee for medical services provided by private practitioners, and approximately 100 million records per year relating to claims are stored and interfaced with a Medicare enrolment file containing personal data. However, as yet, only the information necessary for payment of benefit is being collected: name, date of birth, sex, geographic area, and usual residence, and a provider's identifier. No data are collected about either diagnoses or procedures.

With regard to specific diseases, each state now has a Cancer Registry based on notifications by hospitals and laboratories of cancer patients and by Registrars of deaths attributed to cancer. A proposal to establish a National Cancer Statistics Clearing House is currently before the NH & MRC and, if accepted, national figures should be available within a few years. Instances of communicable diseases are reported to state health authorities and collated by the Department of Health but are of very limited research value. Information on work-related health problems can be obtained from compensation claims to the state-based insurance schemes, but large numbers of workers are not covered and there would be under-reporting because neither medical practitioners nor workers are aware of the significance of work factors in the aetiology of many chronic diseases. Police, legal authorities and traffic

authorities collect information which yields statistics on road traffic accidents, and on drink-driving offences and drug offences. At present, these have no research value, but the Department of Transport is developing a National Road Traffic Accidents data base based on police reports of accidents involving fatalities and casualties.

How can these routine reporting systems be used to develop and check hypotheses about causes of disease? The traditional approach is to produce estimates of the risk of various health outcomes across various subgroups of the population under study. Thus, for example, an indicator of the health risk associated with particular occupations is given by ratios of the number of deaths from various causes to the population at risk and making comparisons across occupational categories. The population at risk is usually obtained from census data. Clearly there are severe limitations with this type of unlinked data analysis, particularly in the lack of ability to apply controls for what might be relevant covariables. This is limited to those factors coded in both the census and health event data, and there are usually very few in the latter - age, sex, location and perhaps occupation, often in groups too broad to be useful.

The value of this type of data is greatly enhanced when it can be linked directly at individual level to data on personal characteristics and lifestyle factors. In this case, files must include full names, previous surnames and dates of birth at least. Concern about confidentiality is, however, very high in Australia and the opportunities for record linkage are very limited, although some useful work has been done by linking cancer records and death records to employment records. However, identifiable census returns have been destroyed in Australia since the start of the century and proposals that a sample of these be retained, as has been done recently in England and Wales, seem unlikely to succeed.

Population Surveys

Given the poor state of routine health data collections, it is fortunate, though perhaps not unrelated, that Australia shows up reasonably well in the area of health surveys in the International Health Data Guide. This is probably due to the preference of the Australian Bureau of Statistics for using its limited resources for population surveys which can cover a range of topics rather than for more specific health related collections. In addition, state health authorities and other agencies have undertaken a number of surveys relating to lifestyle factors, particularly cigarette smoking, alcohol consumption and drug use.

Prevalence surveys of population samples have the advantages that, for established risk factors, such as smoking, they can be used to describe cross-sectional and longitudinal variations which may relate to health outcomes, and they provide a basis for evaluating community behaviour — intervention programmes. For postulated or possible risk factors, they can be used to examine relationships with health outcomes and may provide clues about causal factors in disease.

The Australian Health Surveys in particular provide valuable information on the health status of the population, particularly since recent changes to the ABS Act now permit the release of de-identified unit record data. A data file from the 1977-78 survey has been released, and the 1983 data file should be available later this year. A further survey is planned for 1986.

The surveys give detailed information on a wide range of personal characteristics, although the categories of such important variables as occupation and birthplace were too broad in the unit record file for many research purposes. Categories had been collapsed in order to ensure non-identifiability of respondents,

particularly those in small subgroups of the population. For such subgroups, of course, sample surveys are of little value because of the small number, if any, of cases interviewed. Nevertheless the practice of collapsing variables into standard classifications without giving sufficient thought to the potential research uses needs to be changed.

Interview data on self-reported health status may also be a suspect, and it would be useful to have some medical verification of a subsample of respondents or from pilot tests. Another shortcoming of these data is the lack of information on health risk factors such as smoking and drinking, although 12 items from the General Health Questionnaire were included.

Several community health studies have been carried out in Australia in recent years, and these may provide clues to the relationship between life-style characteristics and health status. One of the advantages of such studies is that they often include clinical checks on the self-reported health status of respondents. The main disadvantage is that the sample sizes are generally too small to allow tests of hypotheses.

The prevalence of heart disease as a major cause of morbidity and mortality has generated a number of studies aimed at determining the associated life-style factors. The best of these are the two Risk Factor Prevalence Studies conducted by the National Heart Foundation in 1980 and 1983. Both studies used large national random samples and included clinical examinations to obtain height, weight, blood pressure and blood lipid levels in addition to interview data on tobacco, alcohol and medication consumption, diet, physical activity and psychological stress.

While major programmes designed to influence the smoking behaviour of the community have been launched throughout Australia, the ABS has chosen to cease conducting surveys on smoking and to exclude questions on smoking

from the Health Surveys. The last national survey on smoking conducted by the ABS was in 1977. A series of much smaller national surveys, on about 6000 adult respondents, have been conducted by the Anti-Cancer Council of Victoria in 1974, 1976, 1980 and 1983. However, a large number of school-based surveys of alcohol, tobacco and drug use have been undertaken, although as a basis for national figures these have problems of comparability. Two national surveys of school children aged 9-16 years were conducted by the NH & MRC in 1969 and 1973, and a similar survey has recently been carried out by the Australian Cancer Society and the National Heart Foundation.

Efforts to reduce the number of road traffic accidents have centred on programmes designed to reduce alcohol use with random breath testing being introduced in most states of Australia and substantial expenditure on television advertising campaigns. However, the availability of statistical data to evaluate the effectiveness of these approaches is limited largely to the mortality and morbidity statistics and some small studies on knowledge, attitudes and behaviour relating to drink-driving.

Concluding Comments

As indicated in the above brief review of Australian health status data, only limited attention has been given to the needs of researchers for aetiological analyses. Routine health collections lack uniformity across state boundaries and do not include sufficient information on parents' background or possibly associated risk factors. Record linkage could overcome some of these deficiencies but has generally been resisted by the appropriate authorities.

Concerns about invasion of privacy, the lack of legislation on the preservation of confidentiality for some collections (morbidity), and over-rigorous interpretation of such legislation in others (census), have restricted the use that could be made of these data. Population surveys have been adopted as an alternative, but these are generally too small for detailed analytic studies and are thus limited to monitoring the prevalence of established risk factors.

The National Health Statistics Workshop held in February expressed its concern over this lack of appropriate data and recommended that a national health statistics agency be established as part of the newly formed Australian Institute of Health. This new agency should promote the development of national collections such as the natural death index, cancer index and morbidity collections, and ensure that the necessary legislation is enacted to provide for preservation of confidentiality. Priority should be given to assembling data already available in most cases at state level into unified national collections, to the development of record linkage procedures, and to risk factor surveys of diet, smoking, alcohol and illegal drugs.

This is obviously a large agendum which will require considerable resources and time to implement. Nevertheless, there is strong support behind the recommendations and a reasonable hope that a substantial improvement in Australian Health Statistics will be achieved. □

Providing Local Data Services

by R. de Vries¹
Steinmetz Archives/SWIDOC.
Amsterdam, The Netherlands

The Steinmetz Archive, as the Dutch national data archive for the social sciences, is a somewhat special case in the context of this workshop. We function both as a data clearinghouse and as a data library per se. As data library, the archive operates in an area that, in other countries, would be seen as a "regional" (as opposed to "national"); the following are some reasons why the Steinmetz can serve as an example of a data library providing local data services.

- dependency on someone else's computer center. I.e., decisions on installation of software packages, policy on how to handle mass storage problems and the safekeeping of magnetic tapes, participation in networks, etc., are all beyond our direct control; we can argue, but have no real influence on such decisions, let alone the means of implementing them on our own.
- dependency on the willingness of research

organizations and their funding bodies to deposit data (survey or otherwise) in the Archive's holdings. Again, we have no financial means with which to buy large datasets, nor the manpower, in the case of published statistical data for example, to generate new data from published sources.

- a small staff (5).
- a strong emphasis on documentation and reference service, aided by a reference database containing study descriptions of every stored dataset. "Documentation" here refers to the original questionnaire, research report, print-outs of frequencies, etc.
- given the national role of the Archive a less than desirable situation, but in the context of "local data services" quite reasonable: we cannot give access to our holdings through a network, nor is the reference database available online. Data exchange is via magnetic tape, and available datasets are brought to the attention of potential users via regular newsletters and a published catalogue. A service that is in my opinion typical of local data services, data exchange on floppy discs, is possible but we have hardly any experience with it as yet.

If one sees a data library as a "local" service backed by a central data clearinghouse or central acquisition and processing centre, then indeed one expects an organisation with a small staff, using external computers and software, concentrating on reference as well as actual dissemination in as friendly a manner as possible, generating subsets, documentation and other special requests. Also, seen geographically, the data library should be within one day's travelling distance for its users, for consultation purposes. Given this definition, the

¹Paper prepared for presentation at the IFDO/IASSIST Conference, Workshop on Data Services, Amsterdam, May 20 - 23, 1985.

Steinmetz archive does serve as an example of a data library.

Are there any other local data services in Holland for the social sciences? For survey data: no. For statistical information at the level of cities or regions: yes. There are several specialised databases owned by government organizations for planning and policy making, and by university departments, for research and training. These are, on the whole, "local" in the sense of being accessible only to their own people, not to outside researchers, either academic or otherwise.

In the second part of this account, I will briefly outline our approach in the following areas: data acquisition, data dissemination, storage and maintenance of data, documentation and reference services.

Data acquisition is accomplished by routinely checking registers of ongoing research, social science periodicals, and reports of finished research. This is facilitated by the Steinmetz Archive's participation in the Social Science Information and Documentation Centre. There is no human network of researchers or fund raisers in the field who could report to the Archive interesting projects or data. (Nor would I would expect the kind of data library that provides local data service to rely on such a network for data acquisition.)

Dissemination. Users in Amsterdam, where the Steinmetz Archive is located, have direct access to the data through the local university-owned computer centre (SARA). From a local terminal, a user can get a copy of a dataset by simply starting a job, that has only one variable: the Steinmetz number given to the particular dataset. Central logging of these jobs and who has started them, is automatically reported to the Archive, thus enabling a monthly overview of this type of usage. Other users receive the data, and often an SPSS setup, on tape. This arrangement makes, of course, no

provision for users without at least access to a minicomputer with a tape drive and statistical package, such as SPSS. For example, we are unable by these means to provide service to schools. Data transfer to such users should be through floppy discs, a service that we have not really started yet.

As an archive, with an obligation to disseminate 10 to 15 year old datasets, it requires that we have strong **Data storage and maintenance** systems. This we achieve with a system of multiple tape backups and a tape refreshing scheme to guarantee that no tape is physically more than three or four years old. All tapes are stored in the computer centre. One might expect a data library per se to be more relaxed in these matters; whatever gets lost can be replaced upon request from a central data organisation, but this is not so in our case. Developments in "laser disc" technology could ease local mass storage problems, and at the same time ensure long term reliability.

How is the user introduced to these masses of carefully preserved data? Through **Documentation and reference services.** The Steinmetz Archive, as mentioned previously, helps users find data suitable to their needs through a catalogue, which is easily, and regularly, produced from a reference database, various indices on microfiche and paper, and through the original documentation produced by the principal investigator. Introductions to the principles of empirical research and the data available for secondary analysis by means of making "teaching packages" available and giving lectures at schools and colleges, are other means to assist users. The Steinmetz does not give lectures but does offer a teaching package together with the relevant data; the package was developed by an outside institution. Data libraries, which should need to put less effort into such activities as acquisitions, processing and maintenance, might profitably put more effort into actively getting users acquainted with computer-assisted analysis, data sources, etc. □

The Development of a Canadian Union List of Machine Readable Data Files (CULDAT)

This article is an abridged version of the final Report, "Pilot Project for the Development of a Canadian Union List of Machine Readable Data Files (CULDAT)," prepared by Edward H. Hanis, Social Science Computing Laboratory, University of Western Ontario for the Machine Readable Archives, Public Archives of Canada.

A survey of Canadian social scientists, undertaken in 1982, indicated that a need existed for an inventory or union list of data files available for secondary analysis. In the mid-seventies, the Data Clearing House for the Social Sciences (DCHSS) had spent considerable time and effort in the development of an automated inventory. The loss of DCHSS, due to lack of funding, unfortunately also involved the physical loss of the magnetic tape which held the descriptions of these files. The results of the 1982 survey indicated strongly that the research community continued to feel that a union list of data files would be a valuable resource. In response to this need, the Machine Readable Archives Division (MRA) [of Public Archives Canada] established a contract with the Social Science Computing Laboratory of the University of Western Ontario to develop an online inventory describing computer files held by Canadian data archives and libraries. The

overall purpose was to develop organizational, technical and informational foundations for maintaining and disseminating a computerized inventory. Specific objectives involved: the establishment of a standard for describing MRDF for entry into the data base; the design and implementation of the pilot data base containing a partial inventory; and the definition of the organizational roles and mechanism to effect routine and cost-effective flow of descriptive information from data archives and other organizations to the union list beyond the conclusion of the pilot.

The pilot project was carried out over a fourteen-month period. In January of 1985, a committee of data archivists and data librarians established a list of elements which were to be used to describe the holdings of the institutions. These elements were taken from those defined in the MARC format for data files. A data dictionary was developed to aid participants in the entry of descriptive information. The Social Science Computing Laboratory was involved in six major activities: the creation of the pilot data base; the set-up of online access with Basis on the lab's VAX11/785; the set-up of DATAPAC and standard dial-up communications; conducting an evaluation of the online system; a survey of potential contributors; and production of a hard copy reference document.

Contributors to the data base were from the university-based archives and libraries and included: Data Library, University of British Columbia; the Institute for Social Research, York University; Data Resources Library, University of Western Ontario; Institute for Social and Economic Research, University of Manitoba. The MRA also contributed descriptive entries. In all, 753 records were entered into CULDAT. Evaluation of the data base was extended to more participants than those listed above and included both frequent users of online systems as well as infrequent users. Although a number of suggestions have

been made as to how to improve the online inventory, the general consensus was that the data base was very useful and should be continued.

It is not surprising that the most crucial component of the data base was the description of the data file. A number of difficulties were experienced with the lack of consistent terminology used and the detail of the description itself. The problems encountered are summarized in the following paragraphs. The resolution of these difficulties have formed the basis of the CULDAT work plan for 1986-87.

The choice of data elements to be included in CULDAT was based on the fields in the MARC format for data files. A limited number of elements was chosen as it was felt by the committee that the intention of the data base was to include only sufficient information to identify a unique data file, to aid researchers in selecting files of interest, and to locate archived copies of the file. The resulting CULDAT Data Element Dictionary contained the field names and a brief description. During the pilot project, it was noted that in some cases the data dictionary did not provide sufficient guidance to the archivist or librarian to allow him to adequately describe data files, and presumed a knowledge of the MARC format and Anglo-American Cataloguing Rules II. This created some difficulty in mapping out the information received for input into CULDAT. The consequences of a weak data element dictionary are inconsistent presentation of the information which can make the descriptions difficult for the end user to interpret. Weak data descriptions yield inefficient indexes, which, in turn, require that the user anticipate all possible variations of a term in order to find all relevant records in the data base. Specific problems were found in the following data elements.

1. Investigators: The differentiation

between principal investigator and other investigators caused some difficulties for both cataloguers and the users. The determination of principal investigator for a data file is difficult, if not impossible, at times. The separation of these fields requires searching two fields rather than one for the user wishing to browse the index. The distinction between investigator (personal) and investigator (corporate) was considered essential. The lack of authority control in the corporate investigator field was a problem which could be overcome through the use of Canadiana to control the use and spelling of names.

2. Producer, Generator, Distributor: A tendency to repeat the same data in these fields was found. This may have been due to the inadequacy of the data dictionary. Abbreviations and acronyms were used. The adoption of an authority file for corporate names would apply to these fields as well.
3. File Size; Number of Cases: Some difficulty was experienced in the data provided in this field. Again, this was due to lack of guidance in the data dictionary.
4. Access Restrictions: As all institutions have their own access regulations, it was felt that this field should only be completed when the distributing organization has contributed the record.
5. Abstracts: Information contained in this field was found at times to repeat information found in other fields. The vocabulary used varied widely which made control of the field extremely difficult. The types of variables used in a data file is vital information for the prospective user. In order to provide improved access to this field, it would be

preferable to separate the abstract from the variable list. Variables could then be left unindexed. Such a change would significantly reduce the indexing overhead and improve the quality of the printed keyword index by using variable names instead of individual words. The online system could continue to index variables as individual words as well as expressions.

6. Geographic Coverage: The pattern adopted by the pilot was as follows: site, city, region, territory, province, state, country (qualifier) continent. The pattern worked well in most cases and ensured that the user interested in data about a particular province could retrieve information on a file which covered only a city in that province. The only records which do not conform to this pattern are physical data where orbital coordinates are submitted.
7. Chronological Coverage: The format of the dates recorded in this field was inconsistent, rendering the retrieval of data ineffective. The data dictionary should prescribe one acceptable format to which all dates would be converted. A standard format will provide the possibility of performing systematic retrieval on time periods by scanning the text, even though every unit of time within a range is not actually recorded in the field.

The difficulties which have been encountered will provide valuable information to allow us to improve the quality and guidance required for the data dictionary. The second version should improve the consistency of the descriptive entries. The contributions made by the data archives and libraries were extremely useful in building the pilot data base and allowing us to identify specific areas for improvement in the data dictionary.

User Evaluation and Potential Contributors

The original project design called for online testing and evaluation of the pilot CULDAT data base by project participants and constructing a list and contacting potential contributing organizations in order to learn about their holdings and interest in submitting entries into CULDAT in the future. Three important additions were made to enhance the project. The establishment of a DATAPAC Service reduced usage costs and significantly improved convenience to remote users. In addition, the survey of contributors was expanded to include questions on evaluation as they were potential users as well. The third activity was to include three local University of Western Ontario groups (students in the School of Library and Information Science, the University's reference librarians, and social science researchers who use the Lab's support services). These additions increased the use of CULDAT during the pilot phase.

The evaluation of the data base was very favourable and many respondents expected to benefit from the availability of CULDAT in the future. Considerable information from prospective contributors and users was acquired. This information and experience provide a sound foundation for the design and planning of the next stages in the development of CULDAT. The major activities planned for 1986/87 will include: 1) the revision and expansion of the CULDAT Data Dictionary in order to provide more guidance on the description of holdings for entry into CULDAT;

2) continued support to university based data archives and libraries to ensure their holdings are included in the inventory; and 3) the redesign of the formatted hardcopy version to make it available as a reference document at less cost.□

SES Archiving Policy Responsibilities of Program Officers

SES Policy

Official award letters (from DGC) will now include the following paragraph as a condition of the grant:

All data sets produced with the assistance of this award shall be archived at a data library approved by the cognizant program officer, no later than one year after the expiration date of the grant. In cases that involve issues of confidentiality or privacy, precautions consonant with human subjects guidelines shall be observed.

Program Implementation

The program officer should discuss this requirement with each new grantee for whom it is pertinent, explaining that any data set generated as a product of an NSF award must be archived at some readily accessible, ongoing facility. (In most circumstances, this would be ICPSR, but a grantee who can make a convincing case for archiving the data elsewhere should be free to do so.) The program officer should emphasize that the data to be archived must be fully documented and cleaned and that this must be accomplished no later than one year after the grant period. The grantee's acceptance of this condition and specific plan for fulfilling it must be received, in writing, before the grant is made. Along with the congratulation letter, the grantee will receive informational material to help in archiving the data.

In discussions with grantees, program officers are encouraged to explain the scientific rationale underlying this policy. Program officers should view these discussions as an opportunity to communicate the importance of data access for purposes of replication, verification, secondary analysis, etc. Further, investigators might appreciate understanding that this policy permits leveraging scarce resources for the social sciences insofar as it should virtually eliminate the need for overlapping or redundant data assembly and collection in the future. Also program officers may want to emphasize that the one-year time span (after the grant period) is intended to assure that the grantees have adequate opportunity to make significant progress on their own projects.

Dissemination

This archiving policy and the rationale for it are to be communicated to the scientific community through individual discussions and via association newsletters, sessions at scholarly meetings, workshops, and other gatherings. Program officers are encouraged to play an active role in this process.

Division of Social and Economic Science
National Science Foundation
Washington, D.C. 200550

THE ATTACHED HAS BEEN PREPARED TO AID YOU IN ARCHIVING THE DATA SETS IN YOUR STUDY. IF YOU HAVE QUESTIONS, PLEASE CONTACT THE PROGRAM OFFICER RESPONSIBLE FOR THE ADMINISTRATION OF YOUR GRANT.

Archiving Machine-Readable Data Files:
Preparation Guidelines

The minimal requirements for archiving a machine-readable data set include three basic components: (1) the data file, and (2) a codebook that provides definitions of variables and cases and/or other instructions for interpreting the data file, and (3) format documentation for the medium (e.g., magnetic tape, floppy disk) used to store and transport the data set.

The data file is most frequently a rectangular matrix (rows and columns) of numeric characters and/or alphanumeric strings in a fixed format (i.e., data for any particular variable will be located in the relative position in the record for every observation). The format of the data file should be such that it does not depend on any specific software program in order to be used. In particular, data files should not be supplied in so-called "system" file formats (e.g., SASfiles, or SPSS "Get" files). This includes special formats designed for transportability between different types of hardware installations (e.g., SPSS-X Export files).

The most crucial part of any data set is its documentation. A codebook that clearly defines the variables and cases is essential. The following information should be included for every variable:

1. An unambiguous name or reference number of the item.
2. A textual description of the item, or the text of the question, if from a questionnaire. If the variable is a recode or transformation of other variables in the data file, the exact definition of the recode or transformation is necessary.
3. The starting location, width or ending location, and location of implicit decimal point (if any).
4. Missing data codes and the meanings.
5. The mode in which the variable is represented, i.e., numeric character, alphanumeric string, etc.

The codebook should also contain a list of the valid values for categorical items, and valid ranges for continuous items. Missing data codes should be documented in the same fashion as other values, and not left implicit.

The format of the medium used to store and transport the data set must also be documented. Data are most often supplied on magnetic tape and the following information refers to this particular method.

The number of tracks (7 or 9), density (800, 1600, or 6250 bpi), the character translation set (ASCII or EBCDIC) and labeling scheme (IBM Standard, ANSI, or unlabeled) should be documented. (If a choice is available, data are most reliably stored and retrieved in 9 track, 6250 bpi formats). Data should not be written to tape in formats that require a specific utility program or other software package to read them. In addition to overall tape specifications, the format of each individual file on the tape should be documented. If the tape is labeled, each file label should be indicated together with its particular physical characteristics. These include the record format (i.e., fixed or variable length, blocked or unblocked records), record length (i.e., the number of characters in one record), if blocked, the size of the data block (i.e., number of characters per block or between inter-block gaps, or the number of records per block or between inter-block gaps). Multiple files should be separated by a single end-of-file (EOF) mark. (Multiple EOFs denote the logical end of the tape in some systems).

Data in different or more complex formats than described here may also be archived. The repository in which the data will be submitted should be consulted for the specific data format and documentation requirements.□

Selected References

- David, Martin H., Alice Robbin, Anne Cooper, and Franklin W. Montfort. "Standards for Public Use Files: Lessons from SIPP," June 1985. Madison, WI: Institutue for Poverty Research, University of Wisconsin, Social Science Building, 1180 Observatory Drive, Madison, WI 53706.
- Dodd, Sue. Cataloguing Machine-Readable Data Files. Chicago: American Library Association, 1982. ISBN 0-8389-0365-7.
- Dodd, Sue, and Ann Fox. Cataloguing Micro Computer Files: A Manual of Interpretation for the Anglo-American Cataloguing Rules. Chicago: American Library Association, 1985. ISBN 0-8389-0432-7.
- Geda, Caroline L. Data Preparation Manual. Ann Arbor, MI: Inter-University Consortium for Political and Social Research, 1980. Institute for Social Research, University of Michigan, Ann Arbor, MI 48104.
- Nasatir, David. Data Archives for the Social Sciences: Purposes, Operations, and Problems. New York: UNESCO, 1978. ISBN 92-3-10105-2.
- Robbin, Alice. "Technical Guidelines for Preparing and Documenting Data," in Boruch, Robert F., Paul M. Wortman, and David S. Cordray, Reanalyzing Program Evaluations. San Francisco: Josey-Bass, 1981.
- Roistacher, Richard C. A Style Manual for Machine-Readable Data Files and Their Documentation. Report No. SD-T-3, NJC-62766, Bureau of Justice Statistics, U.S. Department of Justice, June 1980. U.S. Government Printing Office.□

APDU Plans Eleventh Annual Conference

The Association of Public Data Users (APDU) will hold its eleventh Annual Conference at the Ramada Renaissance in Washington, DC, on October 29-31, 1986.

Wednesday afternoon, October 29, will be devoted to a 1990 Census Workshop, utilizing an APDU-prepared paper that summarizes and crystallizes the data product issues. By that time, the U.S. Bureau of the Census will have held its ten Data Product Planning Workshops and will have preliminary plans in place. The APDU Workshop will provide an opportunity to criticize this material and convey informed comments to the Bureau. Thursday morning, the 30th, will start with a keynote speaker followed by discussions on federal information policy. Other sessions that day will include updates on major data series and a look at new forms of information dissemination. Among the sessions planned for Friday, the 31st, are panels on private sector use of public data, population and economic projections, and microcomputers, including demonstrations.

APDU was organized in 1976 to facilitate the utilization of public data through sharing of knowledge about files and applicable software, exchange of documentation, and joint purchasing of data. APDU is committed to increasing the knowledge of its members about new sources of information and increasing the awareness of federal agencies about the requirements of data users.

Program and registration materials for APDU86 will be available in August. For further information, contact Susan Anderson, APDU, 87 Prospect Avenue, Princeton, NJ 08544, (609)452-6025 between 9:30 AM and 2:30 PM.

The United States Information Agency's Visitor Program

In fiscal year 1985, 4,712 visitors from abroad participated in the United States Information Agency's International Visitor Program: 1,941 of these visitors came to the United States at their own or their government's expense, while the remaining 2,771 visitors were fully or partially funded by USIA. The Agency's Bureau of Educational and Cultural Affairs, operating under authority of Mutual Educational and Cultural Exchange Act of 1961 (Fulbright-Hays Act), stimulates and facilitates mutual understanding and cooperation through governmental and private international education and cultural activities.

The emphasis of the International Visitor Program is on communication between people. The program works to strengthen and improve mutual understanding through direct, people-to-people contacts between current and emerging leaders of foreign nations and the people of the United States. Through this program, foreign visitors gain in-depth perceptions of America, and Americans, in turn, learn about the intellectual and cultural diversity of other nations.

Participants in the program are established or potential foreign leaders in government, politics, media, education, science, labor relations, and other key fields. They are selected by USIA and United States embassies overseas to visit the United States to meet and confer with their colleagues and to have in-depth exposure to this country, its culture and people. Over the years, hundreds of former International Visitors have risen to important positions in their countries. As of November 1985, 41 current heads-of-state and 690 cabinet level ministers around the world have participated in educational and cultural exchange programs sponsored by USIA.

The program depends upon the commitment and skills of volunteer-assisted community organizations across the country whose members provide a variety of services, including professional programs and home hospitality, for these distinguished guests. More than ninety of these local organizations are members of the National Council for International Visitors (NCIV) which encourages and promotes efforts to develop, coordinate, and improve services for visitors from abroad. Thus, thousands of Americans across the land contribute to improved international relations through their involvement in the International Visitor Program.

November 1985

Dominion Archivist Celebrates Unique New Course

TORONTO - Dominion Archivist, Dr. Jean-Pierre Wallot, visited Toronto's George Brown College on March 7, 1986 to celebrate the introduction of a unique new course, Machine Readable Records and Archives, to the College's new part-time certificate program in archival practices sponsored by the Toronto Area Archivists Group Education Foundation.

The course is designed to provide students with a knowledge of the techniques required in the management of machine readable information from both the archival and records management perspectives. Based upon an understanding of the methods for inventorying, scheduling and appraising data in a range of automated systems, the course explains how data of archival value could be acquired, processed, described, conserved and made available to the research community. The objective of the course is to enable students to introduce methods and techniques necessary for the integration of machine readable records components in their archival or records management programs. In addition, the course instructors were interested in using the course as a means of developing and testing a training program that could be applied on a broader scale.

Dr. Wallot and members of his staff from the Machine Readable Archives Division who developed the course and provided the instructional services over its seven-week duration were welcomed by J.T.A. Wilson, Dean of George Brown College's Business Division on behalf of the President and Boards of Governors. Also in attendance was John Hardy, the College's Archivist, representing the Toronto Area Archivists Group Education Foundation and acting as master of ceremonies.

In his introductory remarks, Mr. Hardy spoke of the significance of the course to the training of archivists and records managers. "Until now", Mr. Hardy said, "our training programs have continued the traditional emphasis on paper-based systems ... as, increasingly, emphasis in the office is being placed on automated records systems and the maintenance of information in the machine readable format." The course was designed to meet the needs of a growing army of archivists and records managers seeking to establish an integrated approach to the management of all recorded information, regardless of physical form, in their organizations. The course is a reflection of the commitment of the TAAG Education Foundation, the Public Archives of Canada and George Brown College to archives and records management education.

In his remarks, Dr. Wallot, reinforced the significance of the course by reminding the audience that not only was this the first course of its kind in Canada but, indeed, was the first in the world. Noting that because "there are very few organizations in this country that do not use a computer (whether a mainframe, a mini or a micro) to assist in the undertaking of various tasks", Dr. Wallot expressed his concern that future researchers may find it extremely difficult, if not impossible, to undertake their studies if archivists and records managers do not get control over this machine readable information. We could face a large gap in our documentary heritage. "Fortunately", he said, "courses of this nature will expose information professionals to the importance and value of the 'Canadian electronic cultural heritage'."

Both Dr. Wallot and Mr. Hardy spoke of the future - Mr. Hardy looking forward to seeing other such courses being introduced across the land, and Dr. Wallot looking forward to the development of a course workbook or manual that could be used to support this and other courses.

Course instructors included, Harold Naugler, Director, John McDonald, Chief, EDP Information Systems Section, Katherine Gavrel, Chief, Documentation and Public Service, Halyna Kis, Chief, Social, Economic and Cultural Section, Machine Readable Archives Division, Public Archives of Canada.

In attendance were members of the first class of the Machine Readable Archives course, members of the Executive Committees of both the Toronto Area Archivists Group and the Association of Records Managers and Administrators (Toronto Chapter), representatives of the Ontario Ministry of Colleges and Universities, officials of the College, and members of the TAAG Education Committee.

The reception was sponsored jointly by George Brown College's Business Division and the TAAG Education Foundation.

Planning for a second offering of the course, tentatively scheduled for the January 1987 term, is now underway.

In recognition of the importance of a knowledge of machine readable records management to records professionals, the Toronto Chapter of ARMA is giving serious consideration to making such a course compulsory for its chapter certificate.□

New Directions For HRAF

Melvin Ember, Chairman of the Board of Human Relations Area Files, would like to receive suggestions about new kinds of data banks and other services that could be provided by HRAF. The Board of Directors has authorized the officers of HRAF to begin planning for such activities, and we are interested in the kinds of new data banks that would serve the research and teaching needs of anthropologists. The possibilities include data banks on primate and other animal behaviour, linguistic texts and other linguistic materials, ethnohistorical materials and graphic arts. These new data banks, like the present HRAF ethnographic files, will consist of actual texts and other primary materials that are indexed for rapid retrieval. But unlike the ethnographic files, the new data banks will also be computerized. We are interested in finding out what users might prefer with respect to format—online access, floppy disks, video laser disks, tapes, etc. We are also interested in expanding our HRAFlex publications program for all kinds of descriptive data, e.g., organized field notes, coded data. Write to Melvin Ember, c/o the Human Relations Area Files, PO Box 2054 Yale Station, New Haven, CT 06520.□

IASSIST Constitution

ARTICLE I - NAME

The name of this organization shall be the INTERNATIONAL ASSOCIATION FOR SOCIAL SCIENCE INFORMATION SERVICES AND TECHNOLOGY/ASSOCIATION INTERNATIONALE POUR LES SERVICES ET TECHNIQUES D'INFORMATION EN SCIENCES SOCIALES, hereafter referred to as "IASSIST".

ARTICLE II - HEADQUARTERS

The official headquarters of IASSIST will be located with the Treasurer.

ARTICLE III - OBJECTIVES

All activities of IASSIST will be based upon the following objectives:

- 3.1 To encourage and support the establishment of local and national information centers for social science machine-readable data.
- 3.2 To foster international exchange and dissemination of information regarding substantive and technical developments related to social science machine-readable data.
- 3.3 To coordinate international programs, projects, and general efforts that provide a forum for discussion of issues relating to social science machine-readable data.
- 3.4 To promote the development of standards for social science machine-readable data.
- 3.5 To encourage educational experiences for personnel engaged in work related to these objectives.

ARTICLE IV - ACTIVITIES

To accomplish the objectives of IASSIST, some or all of the following activities may be conducted with the approval of the Administrative Committee on a national or regional basis and the submission of an appropriate report:

4.1 COMMITTEES AND GROUPS

Committees may be established and groups of members organized to undertake specific tasks, to find solutions to specific problems, to develop and compile relevant material for specific projects, and to disseminate information on specific subjects.

4.2 CONFERENCES, WORKSHOPS, SEMINARS, TRAINING SESSIONS

Members may convene organized efforts on any subject consistent with IASSIST objectives.

4.3 PUBLICATIONS

A Newsletter will be published and regularly circulated to all members, as well as to others wishing to subscribe. Other kinds of publications may be produced on occasions.

4.4 COOPERATION WITH OTHER ORGANIZATIONS

Efforts will be made to cooperate with other organizations in joint projects and activities when these are consistent with IASSIST objectives.

4.5 OTHER

Other activities that advance the objectives of IASSIST may be undertaken from time to time.

ARTICLE V – MEMBERSHIP

5.1 The membership shall consist of regular and student members, and shall be open to such persons as are interested in supporting the objectives of IASSIST.

5.2 Membership in IASSIST shall include a subscription to the Newsletter.

5.3 Resignations of any members shall become effective immediately upon receipt by the Treasurer of IASSIST. Resignation shall imply forfeiture of the annual membership fee.

ARTICLE VI – FINANCES

6.1 The fiscal year of IASSIST shall begin 1 January and end 31 December.

6.2 Membership fees for regular and student members shall be paid annually to the Treasurer by 1 March of each fiscal year.

- 6.3 The rate of membership fees may be changed by a two-thirds vote of the members on a mail ballot or during the Business Meeting of the General Assembly. Mail ballots will be undertaken between October and December of any calendar year. The results of such ballots or votes will go into effect on 1 March of the following year. In the event of a vote during the Business Meeting of the General Assembly, the membership will be informed prior to the Business Meeting and proxy ballots will be made available.

ARTICLE VII - GOVERNANCE

7.1 GENERAL ASSEMBLY

IASSIST shall consist of a General Assembly composed of all regular and student members. The General Assembly will be organized by geographic regions. The establishment of a region must be approved by the Administrative Committee.

7.2 FUNCTIONS OF THE GENERAL ASSEMBLY

The General Assembly will establish general policies for IASSIST and elect the members of the Administrative Committee, as well as the officers of the Association. Each region will, in addition, elect its own administrative officer who will be known as the Regional Secretary.

7.3 ADMINISTRATIVE COMMITTEE

The Administrative Committee will be the executive body of IASSIST, and shall be composed of at least 10 members elected by the General Assembly from its membership. The composition of the Administrative Committee will reflect the geographic distribution of the members of IASSIST and will be based on the number of members in each geographic region; the Regional Secretaries; the immediate past-President of IASSIST; the President and Vice-President; and the Treasurer, the Editor, and the Secretary-Archivist, the last three individuals having been appointed by the President with approval of the Administrative Committee.

The elected members of the Administrative Committee, including the Regional Secretaries, will serve a three-year term and may serve no more than three consecutive terms.

7.4 FUNCTIONS OF THE ADMINISTRATIVE COMMITTEE

The Administrative Committee will implement policies, develop future directions, and coordinate activities for IASSIST. The Administrative Committee will organize the General Assembly into geographic regions, determine the number of Administrative Committee members from each geographic region, and call meetings of the General Assembly at least once every year. The Administrative Committee will also establish Committees and Groups as required.

7.5 OFFICERS OF THE ASSOCIATION

The Nominations Committee will propose candidates for the offices of President and Vice-President, to be voted upon by the General Assembly. These officers shall serve a three-year term and may serve no more than three consecutive terms.

7.6 ROLE OF THE OFFICERS

The officers of IASSIST will be responsible for the conduct of business of the ASSOCIATION between meetings of the Administrative Committee.

7.7 EXECUTIVE COMMITTEE

The Executive Committee will consist of the Officers, plus other members of the Administrative Committee as required and designated by the Officers.

ARTICLE VIII - MEETINGS

- 8.1 The annual meeting of the General Assembly shall be held at a time and place chosen by the Administrative Committee.
- 8.2 Special meetings of the General Assembly may be called by the Administrative Committee.
- 8.3 The Secretary shall give notice to the members as to the time and place of the annual meeting or special meeting not less than two months prior to the scheduled meeting.
- 8.4 A quorum shall consist of 40 members.

ARTICLE IX - ELECTIONS

- 9.1 A Nominations and Elections Committee will be appointed by the Administrative Committee.
- 9.2 The Nominations and Elections Committee shall conduct an election in each geographic region for officers of IASSIST, members of the Administrative Committee, and the Regional Secretaries. Members within each designated geographic region shall only be entitled to nominate and vote for the Regional Secretary in their home region. However, all members will be entitled to nominate and vote for the officers of IASSIST and the other members of the Administrative Committee.

In the event that competitive circumstances do not exist for a Regional Secretary may be appointed by the Administrative Committee.

- 9.3 A public call for nominations will be sent out by the Nominations and Elections Committee. Voting will be conducted by mail ballot. Elections will be held every three years.

ARTICLE X - AMENDMENTS

The Constitution of IASSIST may be amended by a two-thirds vote of the members on a mail ballot, such ballots to be undertaken between October and December of any calendar year, the results of such ballots to go into effect at the following year's annual meeting of the General Assembly, provided that:

- 10.1 notice of the proposed amendments shall have been given in writing to the Standing Committee on Constitutional Review with the written support of at least five (5) members in good standing of the ASSOCIATION; and
- 10.2 two months' notice of the proposed amendments is given in writing to all members of the ASSOCIATION prior to the conduct of the mail ballot.

ARTICLE XI - TERMINATION

IASSIST may be dissolved by a majority of the members. All property and funds of IASSIST will be transferred to a branch of UNESCO to be determined by the Administrative Committee.

ARTICLE XII - BY-LAWS

SECTION 1

DUTIES OF THE PRESIDENT

- 12.1 The President shall:
- i. be the principal officer of IASSIST;
 - ii. provide leadership and guidance in the realization of IASSIST's objectives;
 - iii. preside at all meetings of the General Assembly and the Administrative Committee;
 - iv. be an ex-officio member of all Standing Committees and shall coordinate their activities;

- v. represent IASSIST in its dealings with external bodies and agencies, particularly those at the international level; and
- vi. report on the state of IASSIST at each annual meeting of the General Assembly.

SECTION 2

DUTIES OF THE VICE-PRESIDENT

12.2 The Vice-President shall:

- i. perform the duties and exercise the powers of the President in the absence or disability of the latter;
- ii. assist the President in recommending measures to further the objectives of IASSIST when and as often as requested;
- iii. be an ex-officio member of all Action and Interest Groups and coordinate their activities, and be responsible for proposing the Coordinators to the Administrative Committee and maintaining regular contact with such Action and Interest Groups throughout the year; and
- iv. in the event of the resignation, death, or incapacity of the President, succeed as acting President for the duration of the then President's term.

SECTION 3

DUTIES OF THE REGIONAL SECRETARIES

12.3 The Regional Secretaries shall:

- i. be the primary officers of IASSIST in their respective regions, working closely with the President of IASSIST;
- ii. provide leadership and guidance in the realization of IASSIST's objectives in their respective regions;
- iii. represent IASSIST in its dealings with external bodies and agencies, particularly those at the national level;

- iv. serve as members of the Standing Committee on Membership;
- v. attend all meetings of the General Assembly and the Administrative Committee; and
- vi. work closely with the Program Director of the Annual Meeting when the latter is scheduled in their particular region.

SECTION 4

DUTIES OF APPOINTIVE OFFICIALS

- i. be appointed by the President of IASSIST with the approval of the Administrative Committee.
- ii. attend meetings of the Administrative Committee and meetings of the General Assembly and shall record all facts and minutes of all proceedings in the books kept for that purpose;
- iii. be responsible for the maintenance of IASSIST's records and for its general correspondence;
- iv. be an ex-officio member of the Nominations and Elections Committee to maintain lists of nominees for office and to assist in the preparation and distribution of ballots;
- v. be an ex-officio member of the Standing Committee on Constitutional Review to maintain notices of proposed amendments to the Association's constitution and to assist in the preparation and distribution of ballots;
- vi. give notice of all meetings of the General Assembly and of the Administrative Committee or President.

12.4.2 The Treasurer shall:

- i. be appointed by the President of IASSIST with the approval of the Administrative Committee.
- ii. have the custody of the funds and securities of IASSIST and shall keep full and accurate accounts of receipts and disbursements in books belonging to IASSIST and shall deposit all monies and other valuable effects in the name and to the credit of IASSIST and in such depositories as may be designated by the Administrative Committee from time to time;

- iii. disburse the funds of IASSIST as may be ordered by the Administrative Committee;
- iv. render to the Administrative Committee at its various meetings, or whenever the members of the Administrative Committee may require it, an account of all his/her transactions as Treasurer and of the financial position of IASSIST;
- v. prepare a written report for submission to the General Assembly at its annual meeting;
- vi. provide the Standing Committee on Membership with up-to-date mailing lists of all members in good standing in each of the geographic regions;
- vii. perform such other duties as may from time to time be determined by the Administrative Committee.

12.4.3 The Editor of the Newsletter shall:

- i. be appointed by the President of IASSIST, on the advice of the Standing Committee on Publications and with the consent of the Administrative Committee, for a term of three calendar years which may be renewed;
- ii. serve on the Standing Committee on Publications; and
- iii. be responsible for the regular preparation, publication, and distribution of IASSIST's official Newsletter.

12.4.4 The Program Director of the Annual Meeting shall:

- i. be appointed by the President of IASSIST with the consent of the Administrative Committee;
- ii. set up and organize the next annual meeting following the appointment;
- iii. be responsible for keeping the Administrative Committee regularly informed of all preparations; and
- iv. work closely with the Regional Secretary in the region in which the annual meeting is to be held.

SECTION 5

COMMITTEES

- 12.5.1 The Administrative Committee at the time of the annual meeting of the General Assembly shall appoint and/or confirm Standing Committees and shall appoint and/or confirm Chairpersons of the said Standing Committees.
- 12.5.2 Standing Committees shall advise the Administrative Committee on matters of policy within their particular sphere, and shall have a Chairperson appointed for a three-year term which may be renewed, two members drawn from the regular membership of IASSIST appointed for a three-year term which may be renewed, one member of the Administrative Committee appointed for a three-year term which may be renewed unless representation from the Administrative Committee is already included in the composition of the Standing Committee in another capacity, and such officers as are designed ex-officio members.
- 12.5.3 The Standing Committees of IASSIST are the following:
- i. CONSTITUTIONAL REVIEW COMMITTEE: responsible for receiving proposals for the enacting, amending, and repealing of the by-laws of IASSIST and for preparing revised articles and by-laws for members' approval, as well as for undertaking an annual review of the constitution and by-laws and proposing amendments as it deems appropriate.
 - ii. EDUCATION COMMITTEE: responsible for the development and advancement of professional programs in education and training and for advising the Administration Committee on the criteria for the approval and certification of such programs.
 - iii. MEMBERSHIP COMMITTEE: responsible for recruiting membership in IASSIST and for recommending alterations in the classes of membership and dues. This Committee's membership shall include the Regional Secretaries.
 - iv. NOMINATION AND ELECTIONS COMMITTEE: responsible for receiving nominations for the election of the Administrative Committee, the Regional Secretaries, and the officers of IASSIST, distributing ballots and electoral information according to regulation, tallying the ballots, reporting on the results of the tally, and for recommending alterations in procedures.
 - v. PUBLICATIONS COMMITTEE: responsible for advising the Administrative Committee on general publications program policy and for reviewing manuscripts submitted for publications. This Committee's membership shall also include the Editor of the Newsletter.

SECTION 6

ACTION GROUPS

- 12.6.1 The Administrative Committee, at the time of the annual meeting of the General Assembly, may appoint Action Groups and for every Action Group so appointed a Coordinator shall be named.
- 12.6.2 A minimum of three (3) members of IASSIST may make application to the Administrative Committee for the establishment of an Action Group at least one month prior to the annual meeting of the General Assembly.
- 12.6.3 Action Groups shall be expected to undertake specific tasks, to find solutions to specific problems, or to develop and compile relevant materials for specific projects. The mandate or terms of reference of Action Groups shall be clearly defined, including the resources and time required and the specific nature of the output or product.
- 12.6.4 Action Groups shall report to the Administrative Committee through the Vice-President on matters relating to their particular sphere, and shall have a Coordinator appointed for a one-year term which may be renewed, two or more members of IASSIST appointed for a one-year term which may be renewed, and such officers as are designated ex-officio members.

SECTION 7

INTEREST GROUPS

- 12.7.1 The Administrative Committee, at the time of the annual meeting of the General Assembly, may appoint Interest Groups and for every Interest Group so appointed a Coordinator shall be named.
- 12.7.2 A minimum of five (5) members of IASSIST may make application to the Administrative Committee for the establishment of an Interest Group at least one month prior to the annual meeting of the General Assembly.
- 12.7.3 Interest Groups shall be expected to disseminate information on specific subjects and to serve as a forum of discussion between as well as during annual meetings.
- 12.7.4 Interest Groups shall report to the Administrative Committee through the Vice-President on matters relating to their particular sphere, and shall have a Coordinator appointed for a one-year term which may be renewed, four or more members of IASSIST appointed for a one-year term which may be renewed, and such officers as are designated ex-officio members.

SECTION 8

NOMINATIONS AND ELECTIONS PROCEDURES

Any regular member in good standing is eligible to hold office in IASSIST.

12.8.1 The Administrative Committee and the Officers.

- i. Every three years, commencing in 1984, the Administrative Committee, President and Vice-President shall be elected from a slate of candidates put forward by the Standing Committee on Nominations and Elections.
- ii. During the fall of any election year, any member in good standing may submit in writing to the Nominations and Elections Committee, the names of as many as seven (7) persons for the slate of candidates regardless of the geographic region in which the nominees reside.
- iv. The Nominations and Elections Committee will compile a list of nominees which shall be reviewed by the Administrative Committee and will mail ballots to the membership during the fall/winter of any election year.
- v. All members in good standing, regardless of the geographic region in which they reside, shall be eligible to vote for a limited number of nominees from each geographic region. The number of nominees from each region will be specified on the ballot, based on each region's percentage of the total membership of IASSIST. Voting will take place over a period of one month during any election year, but in one instance will it extend beyond mid-December.
- vi. The results of the election shall be announced by the end of December in every election year. The results shall be published in the first issue of the Newsletter following the election.
- vii. Newly elected members of the Administrative Committee and the Officers shall take office after the annual meeting of the General Assembly following the elections.

12.8.2 The Regional Secretaries

- i. Every three years, commencing in 1984, the Regional Secretaries shall be elected from a slate of candidates put forward by the Standing Committee on Nominations and Elections.

- ii. During the fall of any election year, any member in good standing in a particular geographic region may submit in writing to the Nominations and Elections Committee, the name of a person for Regional Secretary who must reside in the same geographic region as the nominator.
 - iii. A nomination must be accompanied by a written statement from the nominee declaring his/her willingness to stand for election; a statement indicating that the nominee has institutional support to undertake the duties; and an outline of the qualifications of the nominee.
 - iv. The Nominations and Elections committee will compile lists of nominees and mail appropriate ballots to the membership of each geographic region the fall/winter of any election year.
 - v. All members in good standing in each geographic region shall be eligible to vote for the Regional Secretary for that particular geographic region. Voting will take place over a period of one month during any election year, but in no instance will it extend beyond mid-December.
 - vi. The results of the election shall be announced by the end of December in every election year. The results shall be published in the first issue of the Newsletter following the election.
 - vii. Newly elected Regional Secretaries shall take office after the annual meeting of the General Assembly following the elections.
-

Computers and the Social Sciences

Paradigm Press

Box 1057

Osprey FL 33559-1057

Computers and the Social Sciences is a fresh and unique approach to the newest high-quality research and thinking in this vitally important field. If you:

- want to know how computers and related technology affect society and individuals,
- are doing research on computers and how they are used,
- are a teacher or student evaluating the potential of computers for the social sciences,
- are involved in creating or implementing public policy,
- are a social scientist, computer scientist, policy researcher, or planner interested in changes in societal processes or social research methods caused by the use of computer technology, then, *Computers and the Social Sciences* is an important source of information and ideas for you.

This new journal, introduced by Paradigm Press in the Spring of 1985, is designed to supply information regarding the impact of computers on society, and on the relationship between computers and the social sciences. It is a forum for communication among researchers, analysts, teachers and others who are monitoring and observing developments in this rapidly emerging field of study.

As computers are introduced into the home, workplace and school, they become part of the relationship between client and service organizations, government and citizen, student and educational institution. They affect the fundamental nature of the information we use to understand our world and to make decisions at an individual, organizational and institutional level. Our need to understand the implications of this profound change in our basic societal relationships is becoming increasingly important.

Computers and the Social Sciences will explore how various segments of society, such as students, the elderly, women and persons in mid-career, are affected by computer technology. It will examine the effects on information-gathering, research methodology and analytical thinking in the social sciences. It will explore the ways in which computers are creating change in societal dynamics, such as migration patterns, and the existence of subcultures, the welfare of racial minorities, the communications and power of religious groups, and the mobility of certain classes of persons.

For instance, the first issue of CaSS features a summary of research findings on gender differences in computer learning, a conceptual paradigm for research on home computing, a report of research on the political implications of computer use by local governments, an analysis of the bureaucratic meaning of keyboarding, and a paper on how relational database concepts can improve theory construction in the social sciences. Similar papers will appear in succeeding issues.

EDITOR Ronald E. Anderson*Director, Minnesota Center for Social Research, University of Minnesota***ASSOCIATE EDITORS N. John Castellan, Jr.***Department of Psychology, Indiana University***Dean Harper***Department of Sociology, University of Rochester***Robert Kling***Department of Information and Computer Science,
University of California, Irvine***Sherry Turkle***Program Science, Technology and Society,
Massachusetts Institute of Technology***BOOK REVIEW EDITOR Edward Brent***Department of Sociology, University of Missouri-Columbia***ADVISORY COUNCIL****James S. Coleman**, University of Chicago**James Danziger**, University of California, Irvine**James A. Davis**, Harvard University**G.W. Ford**, University of New South Wales**David Garson**, North Carolina State University**Allen Glenn**, University of Minnesota**Thomas M. Guterbock**, National Science Foundation**Richard Hall**, State University of New York, Albany**David Helse**, Indiana University**Phillip Kraft**, Boston University**Kenneth Laudon**, New York University**Martin L. Levin**, Emory University**Roberta Balstad Miller**, National Science Foundation**Abbe Mowshowitz**, New York University**Nicholas Mullins**, Virginia Institute of Technology**Akifumi Oikawa**, Tsukuba University**James Orcutt**, Florida State University**Robert Pearson**, Social Science Research Council**Judith A. Perrolle**, Northeastern University**Charles G. Renfro**, Review of Public Data Use**Robert H. Seldman**, Journal of Educational Computing Research**Merrill Shanks**, University of California, Berkeley**Philip Stone**, Harvard University**John Storey**, Trent Polytechnic, Nottingham**Manfred Thaller**, Max-Planck-Institut für Geschichte, University of Göttingen**Paul Velleman**, Cornell University**SUBMISSION OF ARTICLES**

All submissions are reviewed by Associate Editors, the Advisory Council, or members of a Board of Reviewers, independent experts who may accept or reject articles or suggest revisions.

Articles should be submitted to the Editor:

Ronald E. Anderson*Computers and the Social Sciences, University of Minnesota
Minneapolis, MN 55454, USA.*

Guidelines for preparation are available on request.

AUTHOR'S BENEFITS

Each author and reviewer will receive 25 reprints of the published paper free of charge. Additional reprints can be ordered at cost.

These can be sent by the publisher to a list supplied by each author.

The first author will receive a free one-year subscription to the journal.

**BOOK REVIEWS
AND SOFTWARE REVIEWS**

Current books and software pertaining to the substantive scope of the journal will be regularly reviewed by appropriate specialists.

Books to be reviewed should be sent to:

Edward Brent, Book Review Editor*Computers and the Social Sciences, Department of Sociology
University of Missouri, Columbia MO 65211*

IASSIST

MEMBERSHIP AND SUBSCRIPTION FEES

The International Association for Social Science Information Services and Technology (IASSIST) is a professional association of individuals who are engaged in the acquisition, processing, maintenance, and distribution of machine readable text and/or numeric social science data. The membership includes information systems specialists, data base librarians or administrators, archivists, researchers, programmers, and managers. Their range of interests encompasses hardcopy as well as machine readable data.

Paid-up members enjoy voting rights and receive the IASSIST QUARTERLY. They also benefit from reduced fees for attendance at regional and international conferences sponsored by IASSIST.

Membership fees are:

REGULAR MEMBERSHIP: \$20 per calendar year

STUDENT MEMBERSHIP: \$10 per calendar year

Institutional subscriptions to the QUARTERLY are available, but do not confer voting rights or other membership benefits.

INSTITUTIONAL SUBSCRIPTION: \$35 per calendar year
(includes one volume of
the QUARTERLY)

Ms. Jacqueline McGee
The Rand Corporation
1700 Main Street
Santa Monica, California
U.S.A. 90406
(213) 393-0411

LIBRARY

APR 8 1960

Department of Library Science
University of North Carolina